The expansion of mutant clones in tumorigenesis

Trevor A Graham

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

UCL.

Centre of Mathematics and Physics in the Life Sciences and Experimental Biology

(CoMPLEX)

University College London

July 20, 2009

I, Trevor A Graham, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

For my family

Abstract

The formation of a cancer is an evolutionary process. A somatic cell may acquire an (epi)-mutation that gives it a growth advantage relative to its neighbours. Progression to cancer occurs as cells in the resulting mutant clone acquire additional mutations, which may also confer a selective advantage. This thesis investigated the expansion of mutant clones both prior to, and during, tumour growth.

First, the clonal expansion of a mutant lineage within a stem cell niche was considered. The behaviour of mutant germline stem cells in the Drosophila testis, a remarkably well-characterised niche, was investigated. A model was developed which identified stem cell phenotypes that were selectively advantageous in the niche.

Crypt fission is thought to be the primary mechanism of clonal expansion in the gut. A model of crypt fission was developed, which was used to estimate the frequency of fission events in the normal human colon. Crypts were found to divide infrequently, and fission was predicted to decrease the likelihood that a mutant clone would be lost from the gut.

Next, the expansion of mutant clones during the initial growth of colorectal adenomas was considered. Individual crypts were micro-dissected from small adenomas, and the mutation status of the APC and K-ras genes in each crypt was determined. Combining this information revealed how the adenoma had formed. It was found that K-ras mutations may have occurred earlier in tumorigenesis than had been previously established.

Last, the effect of clonal expansion on the genetic heterogeneity in a cancer was examined. A model of microsatellite slippage during cancer growth was developed. The model showed that clonal expansion during cancer growth, coupled with a low rate

of somatic mutation, can generate non-negligible amounts of genetic diversity. This suggested that there was little evidence for a low-level microsatellite mutator phenotype in colorectal cancers.

Acknowledgements

First of all I would like to thank Ian Tomlinson and Karen Page for their willingness to be my supervisors, and for keeping their faith in me throughout my PhD. I am deeply grateful to Ian Tomlinson for his ideas, guidance, and encouragement which have shaped the majority of this thesis. I would like to say a big thank you to Karen Page for her mathematical input, encouragement and support at difficult times.

I am indebted to the members of the Molecular and Population Genetics laboratory at London Research Institute of Cancer Research UK who have so patiently taught me how to do some molecular genetics. I am particularly grateful to Andrew Rowan, Annie Lewis, Angela Jones, Christina Thirwell and Sarah Spain. I am also very grateful to the members of the Histopathology Laboratory at the London Research Institute who have inspired me, guided me, supplied me with their data and provided much needed amusement. In particular I would like to thank Nicholas Wright for his generous scientific input and guidance, and for allowing me access to his data. I am particularly grateful to my collaborator Simon Leedham whose ideas, energy and enthusiasm has helped to form this thesis. Thank you too Stuart McDonald and Toby Hunt.

I would also like to say thank you to all those in the Equipment Park and Histology Unit at the London Research Institute who have diligently performed histology and sequencing for me.

Thanks are also given to Darryl Shibata for willingly letting a mathematician with no experimental experience loose in his lab, and for (trying) to teach to me to surf. I am very grateful for our discussions, the cajoling and the teasing, they have all shaped this thesis.

Thank you too Lewis Dartnell, Christopher Hornsby and Lisa Willis. You've provided

enthusiasm, encouragement, ideas and welcome distraction throughout this PhD. Additionally, I would like to say a warm thank you to all members of CoMPLEX for providing a stimulating and friendly environment to work in.

I would like to say thank you to all of my family and to my friends for keeping me going over the last few years, especially so in the trying times towards the end. Georgie deserves a special mention - you kept me going with this for so long. My parents deserve extra thanks, especially for the supply of tea and food and encouragement as I struggled to write up.

Contents

List of Figures

List of Tables

Table 1: Frequently used abbreviations.

Item	
Abbreviation	Description
AA	amino acid
AARs	20 amino acid repeats
APC	Adenomatous Polyposis Coli
BrdU	Bromodeoxyuridine
cnn	centrosonin
COX	cytochrome c-oxidase
G6PD	glucose-6-phosphate dehydrogenase
GSC	germline stem cell
GI	gastrointestinal
JAK-STAT	Janus kinase-signal transducer and activator of transcription
K-ras	Kirsten ras
LOH	loss of heterozygosity
MCR	mutator cluster region
MGMT	O6-methylguanine-DNA methyltransferase
MS	microsatellite
MSI	microsatellite instability
MSI-H	high level microsatellite instability
MSI-L	low level microsatellite instability
MS-S	microsatellite stable
MMR	mis-match repair
ODE	ordinary differential equation
PCR	polymerase chain reaction
SSCP	single strand conformational polymorphism
Upd	Unpaired

Chapter 1

Introduction

1.1 Introduction

Cancers originate from somatic cells that have acquired alterations to their DNA. These alterations, either genetic or epigenetic, cause disruption to the cellular machinery that regulates cell growth (HANAHAN and WEINBERG, 2000). The mutated cells can evade normal homoeostatic control and initiate neoplastic growth.

Human bodies consist of many trillions of cells, many of which divide frequently. Cell division requires DNA replication, so that each daughter cell is supplied with a version of the genetic information and is able to produce proteins necessary for cellular function. DNA replication is an imperfect process though, so each time a cell divides, replication errors, or *mutations*, are introduced into the newly produced strand of DNA. Successive rounds of division cause an accumulation of mutations. When viewed in this manner, the eventual acquisition of a tumorigenic mutation appears to be an inevitability. Yet cancers are relatively rare; in Britain for example, where the average life expectancy of a female is around 81 years (UK OFFICE OF NATIONAL STATISTICS, 2004), only about 1 in 3 people will have cancer during their lifetime (CANCER RESEARCH UK, 2008). This is due, in part, to that fact that the human body has evolved antitumorigenic mechanisms that limit the acquisition of mutations and restrict their spread (KOMAROVA, 2005; PEPPER *et al.*, 2007), or also perhaps limit their effects. One such example of an antitumorigenic mechanism is the *hierarchical structure* of tissues, notably the colonic epithelium. In the colon,

the epithelial cell population is maintained by a small number of *stem cells*. The stem cells divide infrequently to produce *differentiated cell* offspring. Subsequent *amplifying divisions* of the differentiated cells produce the large numbers of cells in the epithelium. Differentiated cells have a limited life-span; in the bowel of the mouse, differentiated cells are shed from the epithelium within a week of forming (POTTEN and LOEFFLER, 1990). Differentiated cells therefore have little opportunity to acquire multiple tumorigenic mutations. The potential for tumour growth is thus limited, since mutations may only accumulate in the small number of infrequently dividing stem cells. The first part of this thesis examines how mutations can arise and spread in hierarchical tissue structures.

To transform a normal somatic cell to a cancer cell requires alteration of the cell's phenotype. In their seminal review, Hanahan and Weinberg suggested that a somatic cell had to acquire six phenotypic traits in order to become cancerous (HANAHAN and WEINBERG, 2000). These traits, self-sufficiency in growth signals, the ability to evade anti-growth signals, sustained angiogenesis, evasion of apoptosis, limitless replicative potential and the ability to invade tissue and metastasise, were termed the *Hallmarks of Cancer*. Acquisition of these six phenotypic traits appears to require mutations in multiple genes. Thus, the somatic evolution of a particular cancer can be viewed as the stepwise accumulation of mutations, with each mutation, or combination of mutations, bestowing or permitting the cancer to acquire another Hallmark behaviour. This stepwise model of tumour progression that transforms normal somatic cells into malignant cancer cells, termed the *Somatic Theory of Carcinogenesis* (MCCOMBS, 1930; EDLER and KOPP-SCHNEIDER, 2005) provides an important conceptual framework for studying cancer genetics .

To fully understand how tumours form it is necessary to consider the phenotypic changes induced by each genetic change. To cause tumour growth, a genetic alteration must cause a selectively advantageous alteration to the cell's phenotype, so that the mutant cell clone can expand at a greater rate than other somatic cells. A subsequent mutation acquired by a tumour cell is only likely to become dominant in the population of tumour cells if it provides a selective advantage within the tumour population. This tumour *sub-clone* can expand to the detriment of the other tumour cells. Thus,

tumour progression, the stepwise transformation of normal somatic cells to malignant cancer cells, is punctuated by repeated rounds of mutation followed by waves of clonal expansion. This is an important observation: it is not sufficient to simply describe tumorigenesis as the simple stepwise accumulation of mutations, but rather the acquisition of mutations should be considered in tandem with the resulting selection, and clonal expansion, which results from each (epi)genetic change. Furthermore, tumour growth causes a massive expansion of cell number; for example a cancer of around 1cm^3 in volume contains around 10^9 cells. Each of the cells in the clone has the initial tumorigenic mutations, and so presumably most of the cells in the clone are susceptible to additional tumorigenic mutations that cause progression towards to cancer. In this sense, clonal expansion can drive cancer progression by increasing the number of cells that are potential "targets" for additional mutations (MOOLGAVKAR and LUEBECK, 2003). The second part of this thesis examines the roles of mutation, selection and clonal expansion during tumour growth.

This introduction describes somatic stem cells and the regulatory processes that influence their division. Particular reference is made to stem cells in the epithelium of the colon, as these tissues are studied in detail in this thesis. Next, the process of tumour development in the colon is described. Both the genetic and physiological processes that underlie neoplastic growth are considered, and paradigms used to understand tumorigenesis are discussed. In tandem with describing the biology of stem cells and tumour growth, the complementary role of mathematical modelling for understanding tumorigenesis is considered.

1.2 Stem cells

Adult stem cells, hereafter referred to as *stem cells*, are thought to be present in all renewing tissues in the body. Stem cells are *unspecialised cells*, as they themselves do not perform a specialised physiological function within the body. Stem cells however are *multi-potent*, which means that they have the ability to produce progeny that are specialised to perform a particular physiological function. These specialised cells are termed *differentiated* cells. Furthermore, stem cells are thought to have a unique ability to perform *asymmetric division* which results in one daughter stem cell and one

daughter differentiated cell. Thus, asymmetric division allows a stem cell population to self-renew, whilst still producing differentiated cell progeny.

The number of times a stem cell may divide is not thought to be restricted. This limitless replicative potential, called *immortality*, is unique to stem cells. Differentiated cells are thought to divide only a few times before becoming mitotically inactive. In the murine colon crypt for example, differentiated cells are thought to divide only a small number of times before becoming mitotically inactive (POTTEN and LOEFFLER, 1990; BJERKNES and CHENG, 1999). Since stem cell lineages are immortal, they are likely to be the precursors of cancer cells. Their lineages are long enough to have a reasonable chance of acquiring the multiple mutations required to initiate neoplastic growth (CAIRNS, 2002). Furthermore, the stem cell traits of multipotency and immortality resemble the phenotype of cancer cells, suggesting that stem cells require a less radical alteration of their phenotype to become cancerous than more specialised cells (PIERCE and SPEERS, 1988; HANAHAN and WEINBERG, 2000). Stem cells are therefore of great interest to researchers of oncology.

What is it that confers a stem cell phenotype, or stemness, upon a cell? Stem cell identity could be due either to *intrinsic* or *extrinsic* properties of the cell. An intrinsic definition of stemness suggests that some cells are simply "programmed" as stem cells, perhaps due to the epigenetic state of their DNA, whereas others are programmed to be differentiated cells. The extrinsic definition suggests that stemness is determined by the environment of the putative stem cell. Stem cell identity may be conferred by proximity to a source of a particular extra-cellular ligand that induces stemness (WATT and HOGAN, 2000; WEISSMAN *et al.*, 2001). The locale of such factors defines the *stem cell niche*, which persists upon removal of the stem cells (SPRADLING *et al.*, 2001). Stemness must actually be defined as a combination of both intrinsic and extrinsic factors, since residing in the niche would not confer stemness if the internal makeup of the cell meant it was unable to respond to the *niche signal*.

The testis of the fruit fly, arguably the best characterised of all stem cell niches, is a clear example of how stemness is controlled by both extrinsic and intrinsic factors. Stem cells are clustered around of a hub of non-dividing cells (GÖNCZY and DINARDO, 1996). The hub expresses the ligand *Unpaired* (*Upd*) (TULINA and MATUNIS, 2001;

KIGER *et al.*, 2001), which stimulates the *JAK-STAT* pathway that maintains stem cell identity. Proximity to the hub thus defines stemness; as cells move away from the hub, they receive a decreased niche signal and differentiate. A cells intrinsic ability to respond to the niche signal Upd also determines its stemness: stem cells with loss-of-function mutations in the JAK-STAT pathway are rapidly lost from the niche (TULINA and MATUNIS, 2001; KIGER *et al.*, 2001).

In the human and murine colon, stem cells are thought to reside at the base of colonic crypts, adjacent to the pericryptal myofibroblasts (HUMPHRIES and WRIGHT, 2008) (see Figure 1.2). The myofibroblasts secrete some members of the *Wnt* family of growth factors (GREGORIEFF *et al.*, 2005), that are thought to maintain stemness. The colonic crypt is discussed in detail in the following section.

1.3 The Colonic Crypt

1.3.1 Colon anatomy

The colon is the final section of the gastrointestinal tract. The colon follows the small intestine and ends at the rectum. The lining of the colon, the *epithelium*, consists of a mono-layer of cells, pitted with millions of finger-like invaginations called *crypts*. Crypts are the basic functional unit of the colonic epithelium and are responsible for producing the bulk of the epithelial cell population. Colon cancers are likely to originate from the crypts.

1.3.2 Crypt histology

In humans, each crypt contains around 2,000 cells. The population of each crypt is rapidly renewing; cells flow from the base of the crypt, along the crypt axis, and are eventually lost through the lumen. This polarised flow of cells from the crypt base suggests that stem cells, the precursors of all other cell types, are located at the crypt base. Stem cells are thought to produce semi-differentiated cells, named *transit amplifying* (TA) cells (BJERKNES and CHENG, 1999). Division of the TA cells produces the bulk of the cell population present in the crypt. The differentiated cells in a crypt can be broadly divided into four main types (BRITTAN and WRIGHT, 2004b).

Absorptive cells, the *enterocytes*, form the bulk of the columnar epithelium. Three secretory cell lineages form the remainder of the main cell types: the mucin secreting *Goblet* cells; *endocrine* cells that secrete hormones; and *Paneth* cells that have a more specific secretory role. A Haematoxylin and Eosin (see Section 4.3.2) stained crypt is shown in Figure 1.1. The mono-layer of cells can be clearly seen. A cartoon of a colonic crypt, indicating the locations of the different cell types, is shown in Figure 1.2.

The mechanisms that maintain stemness in the human colon are less well understood than in the fly. The Wnt family of extra-cellular ligands form the likely niche signal. Some Wnts, and some of the Frizzled family of receptors, show differential expression along the crypt axis (GREGORIEFF *et al.*, 2005). Wnts are thought to be produced by the underlying pericryptal myofibroblasts, which surround the crypt base. The myofibroblasts, in combination with the stem cells themselves, constitute the stem cell niche. Wnt signalling appears to maintain stemness by stimulating proliferation. Wnt initiates a signalling cascade discussed in Section 1.6.1. The Wnt pathway is required for the maintenance of the epithelium: inhibition or deletion of Wnt pathway components results in the loss of the stem cell compartment (FEVR *et al.*, 2007; KORINEK *et al.*, 1998). Additionally, the *Bone Morphogenetic Protein* (BMP) antagonists *gremlin 1* and *gremlin 2*, the *Notch* signalling pathway and the binding of Wnt factors to the *ephrin receptors* (KOSINSKI *et al.*, 2007) may also influence the maintenance of stemness (reviewed in (CROSNIER *et al.*, 2006)). As cells migrate up the crypt axis, away from the Wnt source, they differentiate and eventually stop dividing.

1.3.3 Crypt clonality

A number of studies have shown that colonic crypts are clonal populations, that is, that they are derived from a single founder cell. In-situ hybridisation for the Y chromosome in an XO/XY chimera patient showed no crypts contained both XO and XY cells (NOVELLI *et al.*, 1996). X-chromosome inactivation, due to methylation, has also been used to study crypt clonality in the human large intestine. In females, CpG island methylation occurs at random on one or other of the two X chromosomes early in

Figure 1.1: Micrograph showing normal colonic crypts that have been H&E stained. H&E staining highlights acidic regions, such as the nucleus in red, and basic regions, such as the cytoplasm, in purple. The architecture of the crypts can be clearly seen - the lining of each crypt consists of a mono-layer of cells. Image taken at 10x magnification.

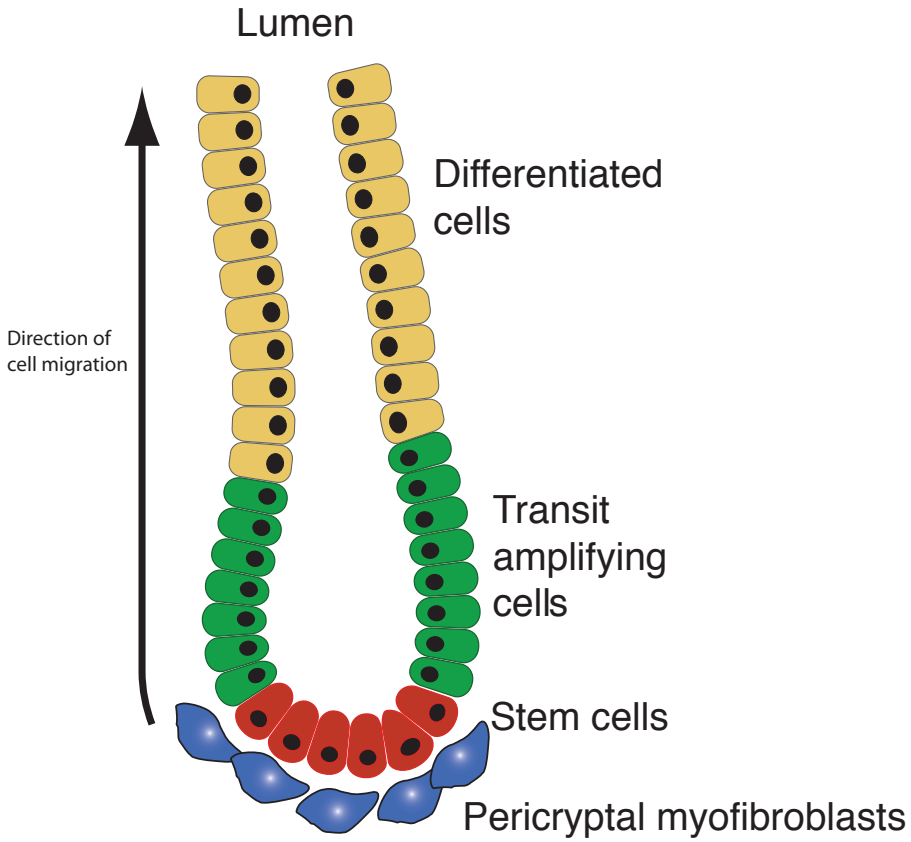


Figure 1.2: Schematic of colon crypt. Stem cells, indicated in red, are thought to reside at the crypt base. Directly above the stem cells are the transit amplifying cells, indicated in green. Terminally differentiated cells are at the crypt top, indicated in yellow. The arrow indicates the direction of cell migration. Below the base of crypt is the location of the pericryptal myofibroblasts.

development, resulting in inactivation of most genes on that chromosome, a process termed *Lyonisation*. Novelli *et al.* examined the colon of Sardinian female patients with a heterozygous functional polymorphism in *glucose-6-phosphate dehydrogenase* (*G6PD*), an X-linked gene NOVELLI *et al.* (2003). During development, X-inactivation led to a patchwork of crypts with either functional or non-functional G6PD expression. Crypt clonality was confirmed by the observation of only entirely wild-type or mutant G6PD expressing crypts, whereas mixed crypts were never seen. Large patches of up to 450 adjacent G6PD deficient crypts were observed, suggesting a common ancestor of adjacent crypts.

Although the population of cells in each colonic crypt undoubtedly has a clonal origin, each crypt contains multiple stem cells. Each stem cell produces its own clone of differentiated cells, so that the population of any crypt is polyclonal in the sense that there are multiple long-lived progenitor cells in the crypt. The first clear evidence that crypts contained multiple stem cells was from experiments using *ethylnitrosourea* (*ENU*) treatment to induce mutations in the intestinal crypts of heterozygous Dlb-1a/Dlb-1b mice (WINTON *et al.*, 1988; WINTON and PONDER, 1990). Loss of the Dlb-1b allele could be detected by an absence of *Dolichos biflorus* agglutinin binding, visualised by standard enzyme histochemistry techniques. Following ENU treatment, many partially staining crypts were observed, whereas after 150 days post treatment, only wholly staining or unstained crypts were observed. Mathematical modelling suggested the data were most consistent with crypts that contained more than one stem cell: ENU treatment induced mutations in only a few of the stem cells in a crypt so that many partially mutated crypts were observed immediately after treatment. After 150 days, these crypts had converted to monoclonality, as a result of one stem cell lineage in the crypt replacing all the others (LOEFFLER *et al.*, 1993, 1997). The mathematical models used to interpret the data are discussed in Section 1.4.4.

1.3.4 Crypt stem cell markers

Hampering the study of colon crypt stem cells, particularly in humans where invasive, perturbation-based experiments are impractical, is the lack of reliable stem cell *markers*. In the colon, numerous molecules have been suggested as stem cell markers.

Of note are *Musashi-1* (*Msh-1*), an RNA binding protein and *Doublecortin and CaM kinase-like-1* (*DCAMBL-1*). Evidence for Msh-1 and DCAMBL-1 as stem cell markers is essentially coincidental: expression of these molecules tends to be restricted to cells found in the putative location of the stem cell niche. However, some authors have noted, for example, that in the human and murine colon Msh-1 expressing cells are occasionally observed away from the crypt base and along the crypt axis (BARKER *et al.*, 2007). Does this mean that stem cells are occasionally found away from the crypt base, or are all Msh-1^+ not stem cells? Clearly, purely associative evidence is insufficient to confirm the efficacy of a putative stem cell marker.

In the colon, a promising candidate stem cell marker has recently been proposed: *leucine-rich-repeat-containing G-protein-coupled receptor 5* (*Lgr5*), a Wnt target gene. In the colon, expression of Lgr5 appears to be restricted to the base of crypts (BARKER *et al.*, 2007). Transgenic mice were used to examine the dynamics of Lgr5 expressing cells (BARKER *et al.*, 2007). In these mice, the progeny of Lgr5$^+$ cells were permanently labelled following induction of a reporter enzyme (LacZ) in these cells. Following induction, labelled cells of all the lineages in a crypt were observed, confirming that Lgr5 expressing cells were multi-potent. Additionally, occasional crypts where all cells were LacZ labelled were observed, proving that Lgr5$^+$ cells were capable of clonal replacement. Lgr5$^+$ cells thus appear to have stem cell like behaviour. It remains to determine the specificity of the marker, as not all Lgr5$^+$ cells are necessarily stem cells . LGr5 may identify stem-like cells, rather than just the actively dividing stem cell population. The discovery of reliable, specific, stem cell markers will permit direct examination of crypt stem cell behaviours.

Stem cell numbers

The lack of a definitive stem cell marker has prevented an accurate count of the number of stem cells per crypt. Estimates of this number rely on interpreting the results of indirect measures, such as methylation pattern analysis in humans (discussed below) (YATABE *et al.*, 2001; NICOLAS *et al.*, 2007). Currently, estimates of the number of stem cells per crypt range from around 4 to over 30 (LEEDHAM *et al.*, 2005; NICOLAS *et al.*, 2007; POTTEN *et al.*, 1997; YATABE *et al.*, 2001). The possibility remains that

differentiated crypt cells retain some *plasticity* and so may be able to de-differentiate to repopulate a depleted stem cell pool. Validation of the stem cell markers noted above may permit definitive and direct determination of stem cell number.

1.4 Crypt dynamics

To study stem cell dynamics, particularly in human colon crypts where direct observation is impractical, a number of indirect assays have been developed. The assays all exploit the power of random infrequent replication errors, which are inherited at cell division, to surreptitiously record cell ancestry. The pattern of replication errors of two cells is compared: cells with similar patterns of replication errors are likely to be closely related, whereas cells with very dissimilar patterns are unlikely to share a recent common ancestor. Of particular note are *mitochondrial DNA* methods and *methylation* based methods.

1.4.1 Within-crypt dynamics: stem cell behaviour

Like any population in a constrained environment, the multiple stem cells in a niche are in competition with one another for the limited resources present in the niche. In the colonic crypt for example, space at the base of the of crypt, proximal to the putative niche signal, is restricted. Thus, it has been postulated that stem cell populations are subject to *genetic drift* whereby one stem cell clone can replace all the others (YATABE *et al.*, 2001). Clonal expansion of a particular stem cell lineage is through *symmetric division*. Rather than dividing asymmetrically, to produce a daughter stem cell and a daughter differentiated cell, symmetric division results in either two daughter stem cells, or two daughter differentiated cells. Since space in the niche is restricted, symmetric stem cell division resulting in two daughter stem cells must be balanced by another stem cell being displaced from the niche. The latter stem cell clone is then lost from the crypt, whereas the first stem cell clone has expanded.

Methylation analysis

Persuasive evidence for the dynamics turnover of stem cells in colonic crypts as described above was provided by YATABE *et al.* (2001). Yatabe and colleagues used

methylation patterns as clonal markers to study within crypt heterogeneity. Methylation occurs at genomic regions with a high cytosine and guanine content, so called *CpG islands*. Specifically, cytosine bases as the $5'$ position relative to a guanine base (a CG motif) can acquire a methyl group (CH_3). The methylation status of each CG site can be determined by *bisulphite treatment*. Bisulphite converts all unmethylated cytosine bases to uracil bases. The bisulphite converted DNA can then be sequenced in the usual manner, and comparison to the wild-type (unconverted) sequence reveals the methylation status of each CG site. The collective methylation status of all sites in a CpG island, a binary sequence, is described as the *methylation pattern* of the locus. In the colon, and in other tissues, such as the small intestine, hair follicles and endometrium, methylation of non-expressed genes notably, *biglycan preproprotein* (*BGN*), *NK2 transcription factor related, locus 5* (*CSX*) and *myogenic factor 3* (*MYOD1*), appears to occur at random, at a relatively low rate, and be inherited at division (KIM *et al.*, 2004b, 2005a; CHU *et al.*, 2007). Thus random methylation events record a cell's ancestry: two cells that have very similar or identical methylation patterns are likely to be closely related.

Yatabe and colleagues looked at the diversity of methylation patterns present in individual colonic crypts. Individual crypts were disaggregated from fresh colon biopsies, bisulphite treated and PCR amplified. To achieve single cell resolution again, PCR products were cloned. An example of the methylation patterns observed is shown in Figure 1.3. Multiple distinct methylation patterns were observed in single crypts, indicating that the crypt contained multiple stem cells that were independently methylated. Stem cell dynamics were determined from the methylation patterns using a mathematical model. A crypt was assumed to contain N stem cells that divided synchronously. Each stem cell could produce 0, 1 or 2 stem cell offspring, and a corresponding 2, 1 or 0 differentiated cell offspring, with probability p_0, p_1 and p_2 respectively, constrained so that each crypt contained precisely N stem cells after division. The intra-crypt variation in methylation patterns was most consistent with a model where $p_1 < 1$ ($p_1 \simeq 0.95$), suggesting that stem cell lineages were not immortal, and a dynamic process of stem cell clone expansion occurred within a crypt. With $p_1 = 1$ the model predicted more variation in methylation patterns than was observed.

Occasional symmetric stem cell division ($p_1 < 1$) could lead to one stem cell clone displacing another, so that the variation in methylation patterns was reduced.

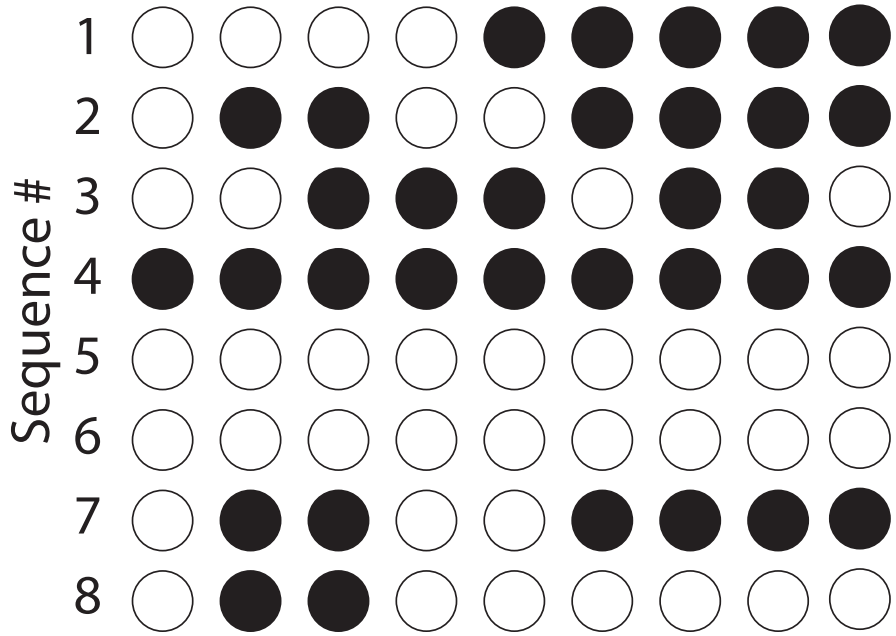


Figure 1.3: Methylation patterns present in a single colon crypt. Each row represents a single methylation pattern, each column the methylation status of each methylation site (black methylation, white unmethylated). Methylation patterns were determined by bisulphite treatment of DNA, followed by PCR amplification and cloning. 6 distinct methylation patterns were present in the crypt, indicating that the crypt contained multiple long lived progenitor cells. The crypt was from a patient aged 75. Data were collected in the Histopathology Section, Cancer Research UK London Research Institute, by the author.

Consecutive rounds of symmetric division can, in principle, lead to *niche succession*, whereby one stem cell clone can expand to displace all the other stem cell clones in the niche. Using their model, with $p_1 = 0.95$ and $p_0 = p_2$ and assuming one stem cell division per day, YATABE *et al.* (2001) calculated a mean niche succession time of 8.2 years. A subsequent more sophisticated analysis of the methylation patterns suggested a longer niche succession time of between 15 and 40 years (NICOLAS *et al.*, 2007).

It was not possible to estimate accurately the number of stem cells per crypt directly from the methylation patterns. A sophisticated analysis suggested a large number of

stem cells per crypt, around 20 (NICOLAS *et al.*, 2007), but stem cell numbers between $N = 2$ to $N \geq 64$ were consistent with the data (YATABE *et al.*, 2001).

Mitochondrial DNA analysis

Mitochondrial DNA (mtDNA) based assays have also been used to great effect to study crypt stem cells. Somatic mtDNA mutations occur randomly as part of the normal ageing process, as mitochondria have poor DNA base excision repair mechanisms, reside in an oxidative environment and have no histones to protect their DNA (MCDONALD *et al.*, 2006a). The mutation rate of mtDNA is sufficiently low, so if two cells share the exact same mutation then they are likely to share a common ancestor. GREAVES *et al.* (2006) used mtDNA mutations that result in cytochrome-c-oxidase (COX) deficiency, an enzyme encoded in the mitochondrial genome, to study clonality. COX-deficient cells can be readily recognised by an enzyme histochemistry technique, which also allows the extent of the clone to be easily visualised. Clonality of two COX-deficient cells can be conclusively confirmed by the presence of the same mtDNA point mutation in both cells, identified by sequencing the entire mtDNA genome.

Greaves *et al.* found the number of entirely COX-deficient crypts increased in colons from patients aged about 40 or more. Laser capture micro-dissection was used to isolate individual cells from a mutated crypt and sequencing confirmed that each cell carried the same mtDNA point mutation (GREAVES *et al.*, 2006). This showed that a single mutated stem cell could repopulate an entire colonic crypt. Partially mutated crypts were occasionally observed, suggesting that crypts contain multiple stem cells. Alternatively, partially mutated crypts could be evidence that a mtDNA mutation had been acquired by a transit amplifying cell, so that only the progeny of the TA cell were COX deficient.

mtDNA mutations illustrate the phenomenon of *monoclonal conversion*. If an expanding stem cell clone happens to be mutated, niche succession of the clone results in a crypt where every cell in the crypt carries the mutation. The mutation is described as *fixed* in the niche.

OAT staining

Patients heterozygous for *O-acetyltransferase* (*OAT*) activity, a sialic acid glycoprotein, have also proved informative to study crypt stem cell dynamics. Routine radiotherapy treatment for colorectal cancer induced loss of wild-type OAT function in some colon crypt cells, which was detectable using a routine histochemical technique (CAMPBELL *et al.*, 1994a,b). In biopsies taken within a few months of radiation therapy, many crypts that showed a mixture of OAT^+ and OAT^- cells (mixed crypts) were seen. One year post-irradiation, the number of mixed crypts fell to the background level observed in biopsies from non-irradiated patients. Thus, the *clonal stabilisation time* of the colon, the average of the time taken by a mutated lineage to become dominant in a crypt or be lost from the crypt, was about one year (CAMPBELL *et al.*, 1996). This study was additional evidence for the dynamic turnover of stem cells in the human colon crypt.

1.4.2 Anti-tumorigenic properties of the niche

Niche succession limits the accumulation of mutations within a niche, so may have evolved as an anti-tumorigenic mechanism. Suppose a stem cell niche contains N stem cells, all of which are equally likely to undergo niche succession and repopulate the entire niche with their progeny. A neutral mutation acquired by one of the stem cells has a $1/N$ chance of becoming fixed in the niche, and so most neutral mutations are lost from the niche. Agent-based modelling by PEPPER *et al.* (2007) compared the robustness of hierarchal structured cell populations that permit niche succession, to self-renewing cell populations that do not permit niche succession. Permitting niche succession limited the rate of somatic evolution and reduced the rate at which faster dividing cells, or cells with a reduced tendency to apoptosis, invaded the niche. Using a differential equation model, KOMAROVA (2005) showed that the niche construction within a colon crypt that permitted competition between stem cell lineages reduced the likelihood of stem cells acquiring multiple mutations in the niche.

If cancers do indeed originate from stem cells (PIERCE and SPEERS, 1988; CAIRNS, 2002), then the initial mutations which lead to neoplastic growth are likely to confer a selective advantage in the stem cell pool. This would increase the likelihood that a

tumorigenic mutation is able to evade the anti-tumorigenic mechanisms in the niche, and be retained in the body. Understanding how mutations can form and spread in a niche is therefore of importance to understanding the initial development of tumours. Evidence that pre-tumour, pro-oncogenic mutations, may alter stem cell dynamics in the niche is provided by the study of the colonic crypts of patients with Familial Adenomatous Polyposis (FAP), who carry a germline mutation in the APC tumour suppressor gene (see Section 1.6.1). In these patients, the diversity of methylation patterns in a crypt was increased (KIM *et al.*, 2004b), which could be explained by an increase in stem cell number or a reduced tendency for symmetric division ($p_1 < 5\%$). The crypts from FAP patients also showed greatly increased fission rates (see Section 1.4.3). Therefore, it appeared that the heterozygous APC mutation had a phenotypic effect on stem cells in the crypt. It is unclear whether a sporadic heterozygous APC mutation would be selectively advantageous within a crypt - but establishing this would provide insight into the genesis of sporadic colorectal adenomas.

1.4.3 Crypt-level dynamics: crypt fission

Crypts themselves are also capable of clonal expansion. The mechanism of clonal expansion is believed to be *crypt fission*, whereby a parent crypt divides to form two daughter crypts by bifurcating from the base upward. The bifurcation presumably splits the stem cell population at the base of the crypt between the new daughter crypts.

mtDNA mutations have provided the strongest evidence for crypt fission. GREAVES *et al.* (2006) observed adjacent COX-deficient crypts, that shared identical mtDNA mutations (see Figure 3.2). Sequencing confirmed that the adjacent crypts shared the same point mutation, indicating that they had a common precursor. The patches must have formed through successive division of the first mutated crypt. The mean number of crypts per patch increased with patient age: by age 80 the mean patch size was around four crypts, suggesting that fission occurs throughout life.

Previously, it had been suggested that crypts that were apparently in fission may actually have been in *fusion*. Using the methylation assay described above, KIM *et al.* (2004b) found that the diversity of methylation patterns within a branching crypt was more similar to the diversity of methylation patterns between unrelated crypts, than to

the diversity within a single crypt. This suggests perhaps that the methylation pattern diversity is generated by the mixing of two independent stem cell populations from the fusing crypts. Alternatively, this could be interpreted as evidence that fission itself takes a long time to complete, so that the methylation patterns in the stem cells of each arm of a bifurcating crypt have sufficient time to diverge. It should be noted though that in this study, only 8 bifurcating crypts were analysed, 7 of which were sourced from the normal appearing colon of a single *Ulcerative Colitis* (*UC*) patient. UC patients show increased incidence of crypts in fission, so may not be representative of normal colonic crypts.

Crypt fission rate

In the 1980s, Bjerknes and Cheng suggested that crypts had a natural life-cycle, that they termed the *crypt cycle* (CHENG *et al.*, 1986a; TOTAFURNO *et al.*, 1987). The duration of the crypt cycle is defined by the duration of time between successive crypt fission events. By counting the frequency of crypts in fission in the human colon and making extrapolations about crypt dynamics that were based on mouse data (see Chapter 3), they estimated that the time taken for the number of crypts in the human colon to double was between 9-18 years (CHENG *et al.*, 1986a). This gave a crypt cycle duration between 13-22 years. Analysis of methylation patterns from pairs of crypts from the colon suggested that crypts that were adjacent appeared as unrelated to each other as those that were more than 15cm apart (KIM *et al.*, 2004b). This is additional evidence perhaps that fission is a rare event, so that the time between successive fission events is sufficient to allow the methylation patterns of adjacent sister crypts to diverge. In Chapter 3 of this thesis, the crypt fission rate in the human colon is estimated directly from data on the frequency of mtDNA mutated crypts.

Initiation of fission

Crypt fission appears to be the primary mode of crypt-level clonal expansion in the colon. Understanding how fission is induced could therefore reveal a mechanism to retard aberrant growth in the colon. Crypts in fission tend be larger than average (TOTAFURNO *et al.*, 1987, 1988; WONG *et al.*, 2002). The mathematical models of Loeffler *et al.* described below (Section 1.4.4), suggest that fluctuations in stem cell

number could stimulate fission (LOEFFLER *et al.*, 1993, 1997). The models assume that the stem cell population is in a constant state of growth, and fission is induced when the number of stem cells per crypt surpasses a threshold value. This formulation was consistent with the formation rate of adjacent mutated crypts in Dlb-1a/Dlb-1b mice following ENU treatment (WINTON *et al.*, 1988). Bio-mechanical mathematical modelling suggests that fission could occur due to *buckling* of the epithelial sheet, induced by increased cell numbers (EDWARDS and CHAPMAN, 2007).

The frequency of crypts in fission also appears to be increased in the normal appearing colon of patients with FAP (CHENG *et al.*, 1986a; WASAN *et al.*, 1998; BJERKNES *et al.*, 1997), suggesting a possible role for APC and the Wnt pathway in the direct regulation of crypt fission events. Alternatively, the heterozygous APC mutation in FAP patients may alter the number of stem cells per crypt (KIM *et al.*, 2004b), and so influence the rate of fission.

1.4.4 Modelling colon crypt dynamics

A number of computational models of the colon crypt have been developed. The models have attempted to infer the biological processes that regulate crypt cell dynamics, such as stem cell number, and the rate of niche succession.

A series of iterations of stochastic crypt models by the Loeffler group are particularly noteworthy (see LOEFFLER *et al.* (1993, 1997)). They created an auto-regulatory model of crypt dynamics where the balance between stem cell symmetric or asymmetric division was regulated by the size of the stem cell pool. Symmetric division (probability $p(S)$, where S is the number of stem cells size) to produce two new stem cells is likely when there are few stem cells and asymmetric division ($q(S)$) is likely when there are many stem cells. Stem cells died at a constant rate r. Crypt division (see Section 1.4.3) occurs where the size of the stem cell pool exceeds a threshold value, and cells are randomly assigned to each of the two new crypts. To test the fitness of their model, the model was compared to the formation of $Dlb{-}1b$ deficient crypts in the murine colon following ENU treatment (see Section 1.3.3). The distribution of crypt sizes observed in the model matched the distribution observed in mice, but the rate of formation of adjacent mutated crypts predicted by the model was too low. Another

weakness of the model is the number of free parameters that cannot be estimated directly from biological data. Nevertheless, the work by Loeffler *et al.* represents the first successful attempt to develop a theoretical model to explain crypt cell behaviour. The modelling provides a convincing argument that crypts contain multiple competing stem cells.

Later revisions by the Loeffler group yielded a *lattice-free* model of crypt dynamics (MEINEKE *et al.*, 2001). In this model, the crypt was modelled as a 2D sheet, with periodic boundary conditions. Cell positions were determined by a physical model that approximated cell adhesive forces and visco-elastic properties by a series of springs connecting the nucleus of a cell with its neighbours. A *Voroni tessellation* was then used to approximate the position of the cell boundaries. This rather sophisticated model was able to reproduce the ribbons of mutated cells that were observed in the crypts of $Dlb-1a/Dlb-1b$ mice following NEU treatment (WINTON and PONDER, 1990). However, the ability of the model to produce predictions about crypt dynamics was again restricted by the number of free parameters in the model which could not be estimated from biological data. Extensive work is still required to produce a model that can reproduce the range of phenomena seen in a colon crypt. Such *Systems Biology*-style modelling attempts are restricted by the availability of biological data from which parameter values can be inferred. In the colon crypt for example, data on cell adhesion, stem cell number, stem cell division rate, transit amplifying cell division rate and mechanisms of cell propulsion along the crypt axis are all lacking. In this thesis, a more reductionist approach is taken. Intentionally simple, hopefully parsimonious, models are developed and used to investigate particular phenomena.

Another model of colon crypt dynamics that is particularly noteworthy is that of NICOLAS *et al.* (2007). Their model attempts to infer crypt properties, such as the stem cell number and niche succession rate from methylation pattern data (described above in Section 1.4.1). Their model is stochastic, and in contrast to previous models of methylation data, their model describes the accumulation of methylation patterns as continuous functions of time. The model is fitted to the methylation data using a *Markov Chain Monte Carlo* technique. The model suggests that crypts contain a relatively large number of stem cells, around $N = 20$, and that niche succession takes

as long as forty years.

1.5 Tumorigenesis

Tumorigenic mutations bestow a selective advantage relative to the normal somatic cells, which leads to aberrant growth. In the colon, how does this growth occur?

1.5.1 Bottom-up growth

The primary mode of aberrant growth in the colon appears to be *bottom-up* (PRESTON *et al.*, 2003). This model of tumorigenesis suggests that a mutant cell first forms in a colonic crypt, presumably in a stem cell. The mutated cell then expands to repopulate the entire crypt, and subsequent expansion of the mutated clone is then through *crypt fission* (see Figure 1.4). Repeated rounds of fission form the large number of dysplastic crypts present in an adenoma.

Preston and colleagues provided evidence of bottom-up spread in sporadic adenomas and in adenomas from FAP patients (PRESTON *et al.*, 2003). In both groups of adenomas, they observed that the nuclear accumulation of β-catenin, a marker of proliferation, was restricted to the bottom portion of dysplastic crypts, suggesting that cell growth occur occurs primarily at the crypt base. Proliferative cells along the entire crypt axis was only observed in more developed adenomas. Furthermore, the proportion of crypts in fission was increased in adenomas (WONG *et al.*, 2002; PRESTON *et al.*, 2003), and also in dysplastic *monocryptal adenomas*, the precursors of larger colorectal adenomas (BJERKNES *et al.*, 1997).

Further evidence for bottom-up growth, using point mutations in the APC gene as clonal markers, is presented in Chapter 5 of this thesis.

1.5.2 Top-down growth

Aberrant crypt growth via *top-down* spread has been suggested by the Vogelstein group. In contrast to the results of Preston *et al.* , Shih and colleagues observed proliferating cells at the tops of crypts dissected from small adenomas, whilst proliferation appeared to decrease towards the crypt base (SHIH *et al.*, 2001). This suggests that the dysplastic

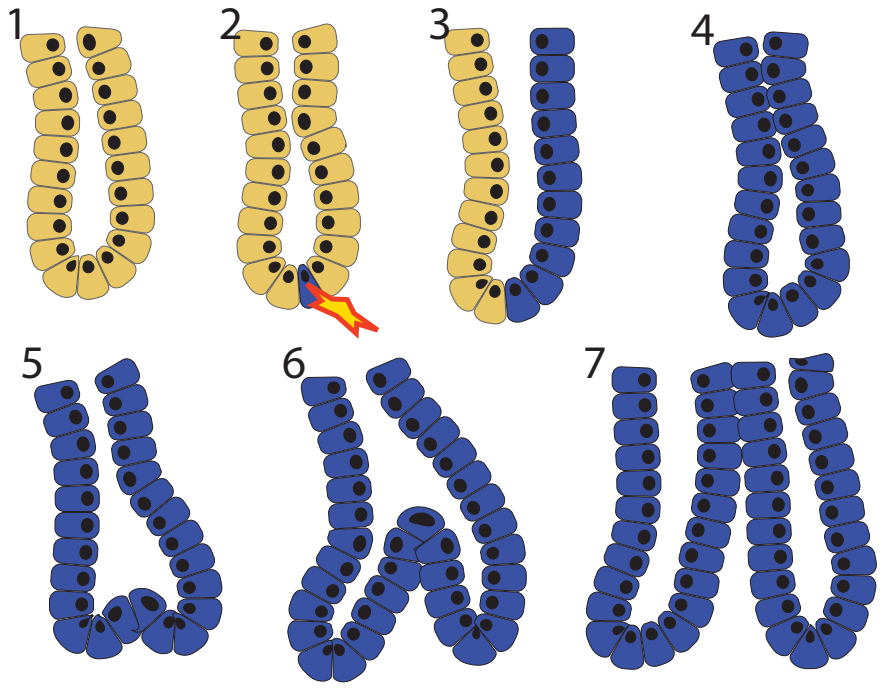


Figure 1.4: Cartoon illustrating bottom up growth in the colon. A mutated cell (coloured blue) forms in the base of a colonic crypt (panels 1-2), and expands to repopulate the entire crypt (panels 3-4). Subsequent expansion of the mutated clone is then through crypt fission (panels 5-8).

cells form, or at least reside, at the top of crypts. Presumably tumour growth is by luminal invasion of normal crypts by the mutant clone.

Top-down spread appears to be generally restricted to later neoplastic growth. PRESTON *et al.* (2003) occasionally observed dysplastic tissue growing down into an adjacent normal appearing crypt. No evidence for top-down spread was observed in very small adenomas or *mono-cryptal* adenomas. In the same study, in-situ hybridisation for the Y chromosome in the adenomas of an XO/XY chimeric patient, who also had FAP, showed that some adenomas were composed of a mixture of crypts that were either XO or XY. However, no crypts were observed that were a mixture of both XO and XY cells, suggesting that mixing of the crypt populations, through top-down spread, was a very rare event.

1.5.3 Field cancerisation

The *field cancerisation* theory proposes that a large patch, or *field*, of epithelial cells may be mutated by exposure to a carcinogen (SLAUGHTER *et al.*, 1953). The field is then predisposed to subsequent neoplastic development. Thus, field cancerisation is a form of *pre-tumour progression*, whereby mutations that contribute to carcinogenesis can accumulate in a large number of cells prior to any overt tumour growth. Evidence for field cancerisation is seen in the colon. In the colon, colorectal tumours with methylation of the O6-methylguanine-DNA methyltransferase (*MGMT*), a DNA repair enzyme, were often found in a field of histologically normal epithelium that also showed MGMT methylation (SHEN *et al.*, 2005), although the role of MGMT in tumorigenesis is unclear. In the colons of patients with *ulcerative colitis*, a chronic inflammatory disease, DNA fingerprinting methods revealed large regions that were clonal for somatic genetic changes (CHEN *et al.*, 2005).

Expansion of the mutated clone, which is presumably through crypt fission in the colon, can increase the size of the mutated field (BRAAKHUIS *et al.*, 2003). Thus it is arguable that crypt fission mediated growth is responsible both for the expansion of pre-malignant lesions, and for early adenomatous growth. Increasing the rate of crypt fission may therefore by a crucial early event in tumorigenesis. This is discussed in Chapter 3.

1.6 Genetic alterations in the development of colorectal cancers

The genetics of colon cancer is probably the best understood of all cancer types. Much of this knowledge has been derived from careful study of patients that have a family history of cancer (FEARON and VOGELSTEIN, 1990; KINZLER and VOGELSTEIN, 1996). Neoplastic growth is usually initiated by mutation the *Adenomatous Polyposis Coli* (APC) tumour suppressor gene (LAMLUM *et al.*, 1999; KINZLER and VOGELSTEIN, 1996). Subsequent changes involve mutations to the *Kirsten-Ras* (*K-ras*) oncogene (VOGELSTEIN *et al.*, 1988), to the p53 tumour suppressor (BAKER *et al.*, 1989), and loss of the *18q* chromosome arm which is the location of the tumour suppressor genes *Deleted in Colon Cancer* (*DCC*) and *SMAD4* (*mothers against decapentaplegic homolog 4*) (VOGELSTEIN *et al.*, 1988). Many other changes may also be important in progression, such as loss chromosome arm *1p* and loss on chromosome 8 (reviewed in (HOULSTON, 2001)).

Attempts have been made to relate particular genetic changes with histological variables that describe the state of the neoplasm. Tumour growth begins by the formation of a small *adenoma*. APC mutations are extremely common in these early neoplasms, whereas other alterations are much less frequent (LAMLUM *et al.*, 2000; KINZLER and VOGELSTEIN, 1996), suggesting that APC is the *gatekeeper* of initial neoplastic growth in the colon (KINZLER and VOGELSTEIN, 1997). Adenomas increase in size and tend to be benign. Eventually, the adenoma may become invasive and is termed an *adenocarcinoma*, or just a *carcinoma*. Invasiveness presumably results from the acquisition of a particular genetic change, or set of changes. Rather than being restricted to the colonic epithelium, adenocarcinomas can grow, or *invade*, the underlying stroma. Finally, cells from the adenocarcinomas may acquire the ability to *metastasise*, presumably through the acquisition of additional genetic changes. Metastatic tumour cells can leave the original *primary tumour* and spread to other parts of the body via the blood or lymphatic system. The transformation of normal tissue to invasive carcinoma, observable as a series of histological changes in the appearance of the neoplasm, is termed the *adenoma-carcinoma sequence* (FEARON

and VOGELSTEIN, 1990; KINZLER and VOGELSTEIN, 1996).

In the following sections, these genetic alterations are discussed in more detail. Particular attention is given to APC, K-ras, and the genetic alterations present in cancers that show *microsatellite instability*. APC is thought to be the first genetic change that occurs in neoplastic development in the colon. The role of APC in tumorigenesis is referred to throughout this thesis. K-ras is mutated in around 40% of colon cancers (ANDREYEV *et al.*, 2001; JONES *et al.*, 2008a). K-ras mutations are associated with an increase in tumour size (VOGELSTEIN *et al.*, 1988). Clonal expansion of APC and K-ras mutant clones in small adenomas is considered in Chapter 5 of this thesis. Microsatellite instability is a type of genetic instability that is associated with colon cancer development. Microsatellite instability is investigated in Chapter 6 of this thesis.

1.6.1 APC

In the late 1980s, genetic linkage studies showed that colorectal cancer patients with Familial Adenomatous Polyposis (*FAP*), a disease of the bowel characterised by the presence of numerous *polyps*, tended to have a heterozygous alteration on the distal arm of chromosome 5 (*5q*) (LEPPERT *et al.*, 1987; BODMER *et al.*, 1987). Subsequently, Adenomatous Polyposis Coli (APC), was implicated as the gene on chromosome 5q responsible for tumour growth (GRODEN *et al.*, 1991). APC is a tumour suppressor, and both allelic copies of the gene must be mutated for neoplastic growth to occur. FAP patients are thus much more likely to develop cancer, as they already have one defective copy of APC, so only a single somatic mutation, or *hit*, on the second allele is required for tumour growth.

APC mutations are found in upwards of two thirds of all colorectal cancers (MIYOSHI *et al.*, 1992). However, the true mutation frequency may be much higher, as the large size of the APC gene makes screening more than a small region of the gene for mutations technically difficult. In most colon cancers, if not all, mutation of APC results in the initial tumour growth (LAMLUM *et al.*, 2000).

Wnt-signalling

The tumour suppressing behaviour of APC is certainly to do with its involvement in the canonical *Wnt* signalling pathway (see Figure 1.5). Wnts, the human homologs of the Drosophila *Wingless* proteins, are a family of secreted proteins that stimulate cell proliferation. In the absence of Wnt, in the cytoplasm of the cell, APC forms the so called *destruction complex*, along with *Axin* and *Glycogen Synthase Kinase 3*β (*GSK-3*β). The destruction complex recruits the intra-cellular signalling molecule β*-catenin*. Whilst bound to the destruction complex, β-catenin is phosphorylated, marking it for subsequent degradation.

The Wnt ligands bind to the *Frizzled* family of cell surface receptors. The Frizzled receptors usually interact with adjacent LRP transmembrame proteins. Binding of Wnts to the Frizzled receptors results in the recruitment of the protein *Dishevelled* to the plasma membrane. In turn, this results in the recruitment of Axin, and possibly the entire destruction complex, to the membrane, causing inhibition of the destruction complex. Thus, β-catenin is no longer degraded and so can accumulate in the nucleus, where it stimulates proliferation through binding of transcription factors. For a detailed review of canonical Wnt signalling see KLAUS and BIRCHMEIER (2008).

Mutation locations

Most APC mutations are found between codons 1250-1450, and this region is correspondingly termed the *mutation cluster region* (MIYOSHI *et al.*, 1992). Furthermore, most mutations in APC that are observed in colon cancers result in a stable truncated protein product. The truncated protein appears necessary for cell proliferation: in cancer cells that express only a truncated APC protein, inhibition of the protein product reduces cell proliferation (SCHNEIKERT and BEHRENS, 2006). Knudson's famous two-hit hypothesis states that tumour suppressor genes require two inactivating mutations, one on each allele, to permit tumour growth (KNUDSON, 1971, 2001). It appears that APC is not a *classical* tumour suppresor gene in this sense, since removal of all the APC protein is very rare, probably because this does not provide an optimal growth advantage. Presumably then, the spectrum of APC mutations observed in colon cancers is the spectrum of mutations which provide an

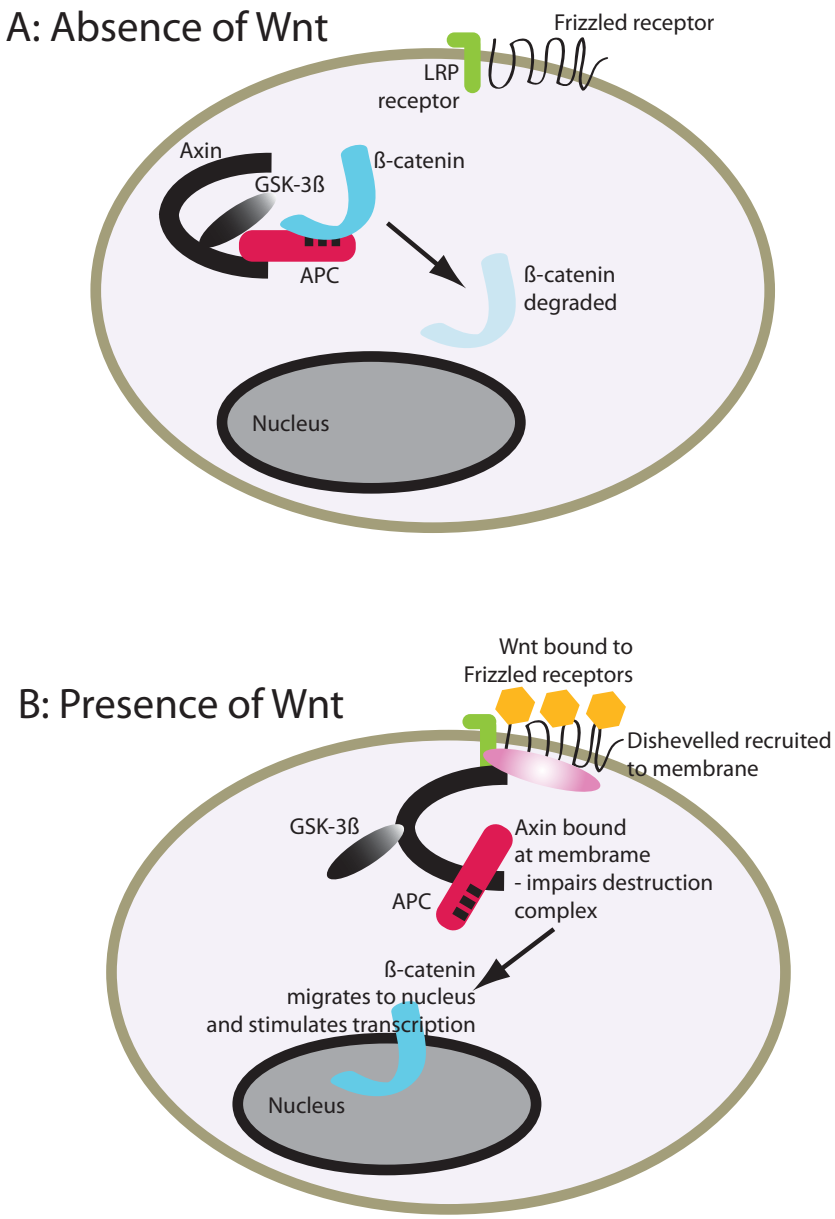


Figure 1.5: Canonical Wnt signalling pathway. **A:** In the absence of Wnt, β-catenin is marked for degradation by the destruction complex involving APC.
B: In the presence of Wnt, Axin is localised to the cell surface, leading to inactivation of the destruction complex. β-catenin is then able to accumulate in the nucleus where it stimulates cell proliferation through transcription factor binding.

optimal growth advantage to tumour cells.

The spectrum of APC mutations selected in cancers must be related to the function of the resulting mutated protein. The middle of APC, at codons 1266, 1379 and 1495 respectively, is the location of the so called 20 amino acid repeats (20 AARs) (SEGDITSAS and TOMLINSON, 2006). The 20AARs are responsible for β-catenin binding. Thus, truncation of the protein between codons 1250-1450 affects the efficiency of APC to bind β-catenin and presumably affects the concentration of β-catenin in the nucleus. The first SAMP (serine alanine methionine proline) repeat in APC is located at codon 1580. SAMP repeats bind Axin, and so are required for the formation of the destruction complex. Nearly all mutations in APC truncate the protein before the first SAMP repeat.

Just-right mutations

The two hits at APC are non-random (SEGDITSAS and TOMLINSON, 2006), in that the location of the mutation on the first allele will tend determine the type or location of the mutation on the second allele. In this sense, the type of first hit selects the second hit. The combination of hits appears to select for the retention of an "optimal" number of 20AARs. The association between first and second hits is most clear in the colonic tumours from FAP patients, where the first "hit" is in the germline (LAMLUM *et al.*, 1999; ALBUQUERQUE *et al.*, 2002; CRABTREE *et al.*, 2003). A germline truncation between the first and second 20AARs is strongly associated with a second hit via loss of heterozygosity (LOH). LOH is usually the result of mitotic recombination. A germline truncation before the first 20AAR is usually associated with a somatic mutation between the second and third 20AAR. A germline mutation after the second 20AAR usually results in a somatic mutation before the first 20AAR. Combinations of mutations in APC thus select for retention of between 1 and 2 20AARs. Similar associations are seen in sporadic colon cancers (ROWAN *et al.*, 2000).

Disease severity

In FAP patients, germline mutations around codon 1309 are associated with the most severe disease (NUGENT *et al.*, 1994). Truncations before the first 20AAR are

associated with less severe disease than mutations in the MCR (CRABTREE *et al.*, 2002). These differences may be due to variations in the efficacy of Wnt signalling in the mutated cells. Mutant APC proteins that retain the first 20AAR β-catenin binding repeat are able to modulate the recruitment of β-catenin to the nucleus (SCHNEIKERT and BEHRENS, 2007), suggesting that the just-right combinations of APC mutations select for an "optimal" level of Wnt-signalling.

Additionally, APC mutations around 1309 seem to require LOH, a relatively more frequent event than a truncating mutation, to achieve a 'just-right' combination of mutations. Thus, the rate of polyp formation in the bowel of an 1309-FAP patient may be higher than in other FAP genotypes, which may explain the increased disease severity.

1.6.2 K-ras

Mutations in the K-ras oncogene are a frequently observed alteration in colorectal cancer, occurring in around 40% of all cancers (WOOD *et al.*, 2007; ANDREYEV *et al.*, 1998). K-ras mutations are associated with an increase in tumour size (VOGELSTEIN *et al.*, 1988). Unlike APC, K-ras mutations tend to be restricted to just two codons, 12 and 13. In a wild-type sequence, codons 12 and 13 both code for glycine. Mutations resulting in a change to a valine or aspartine at codon 12 (G12V and G12D mutations respectively) are most common.

Wild-type K-ras proteins are located against the inner cell membrane layer and maybe either in an active or inactive state. Activation of K-ras can initatiate multiple signalling cascades that can promote cell proliferation, angiogenesis and suppress apoptosis, through pathways including the RAS/RAF, RAS/PI3-K and RAS/RAL (reviewed in CASTAGNOLA and GIARETTI (2005)). Wild-type K-ras remains activated for only a short time before hydrolysis reverts the protein to its inactive state. Oncogenic mutations in K-ras cause a constitutively active protein product, presumably causing sustained activation of pathways that promote cell proliferation and angiogenesis, and suppress apoptosis.

As with mutations in APC, certain K-ras mutations are associated with a worse prognosis. A large meta-analysis showed that K-ras G12V mutation is associated with

the worse clinical outcome than any other mutation type, increasing both the risk of mortality and remission (ANDREYEV *et al.*, 2001).

K-ras is also implicated in Wnt signalling. Transfection of a mutant K-ras G12V allele into cancer cell lines with wild-type K-ras alleles caused an increased in the nuclear accumulation of β-catenin and inhibited the activity of GSK-3β, and appeared to act through the PI3-K pathway (LI *et al.*, 2005). Other in-vitro work on colorectal cancer cell cultures has shown that the over-expression of β-catenin can cause increased activity of downstream targets of the RAS-MEK-ERK pathway, and conversely, over-expression of APC inhibits the activity of the targets (PARK *et al.*, 2006). Thus, the signaling pathways controlled by K-ras and APC may overlap, and so the mutational spectrum of the two genes may be linked. This idea is considered in Chapter 5 of this thesis.

1.7 Genetic instability in tumorigenesis

Tumorigenesis requires alterations to a number of genes. If the rate of mutation per gene is μ per cell division, then, asymptotically, in a static population, the rate at which n genes are mutated is proportional to μ^n (ARMITAGE and DOLL, 1954). If the rate of gene mutation is small, and a number of genes are required for cancer formation, then this rate is vanishingly small. Thus, genetic instability has been suggested as an important driving force in tumorigenesis (RAJAGOPALAN *et al.*, 2003; KOMAROVA *et al.*, 2002). Consequently, many groups have investigated the role of genetic instability in tumorigenesis. Much work has been focused on estimating the proportion of cancers that show genetic instability, and at which stage of tumour development instability arises.

Genetic instability can be broadly classified into two main types: *chromosomal instability* (CIN) and *microsatellite instability* (MSI). Evidence for chromosomal instability is discussed briefly in the following section, but chromosomal instability is not considered in this thesis and so a detailed discussion is beyond the scope of this introduction. Instead the reader is referred to the reviews of RAJAGOPALAN *et al.* (2003) and SIEBER *et al.* (2003). Microsatellite instability is considered in detail in Chapter 6 and correspondingly, is considered in more detail here.

1.7.1 Chromosomal instability

CIN^+ cancers tend to be defined by their having frequent gross chromosomal alterations, usually aneuploidy or polyploidy. In contrast to MSI tumours (discussed below), CIN has not been established as a dynamic process representative of the cell having acquired a mutator phenotype, and, correspondingly, definitive genetic alterations that cause CIN have not yet been established.

A minority of cancers show frequent large scale part-chromosome deletions, regarded as evidence of chromosomal instability (JONES *et al.*, 2007). $Apc^{-/+}$ 1638N mutant mice show evidence of aneuploidy in small adenomas, suggesting that some kind of genetic instability occurs early in tumour development (ALBERICI *et al.*, 2007). Apc is involved in mitotic spindle orientation, and germline defects in Apc cause chromosome misalignment in mouse embryonic stem cells (FODDE *et al.*, 2001). In $Apc^{-/+}$ 1638N mice which also had loss of p53, a more severely disrupted karyotype was observed, suggesting chromosomal instability may be a more prominent feature of later cancer development (ALBERICI *et al.*, 2007). On the other hand, SIEBER *et al.* (2002) found little evidence for CIN in 55 human colorectal adenomas: most adenomas were diploid and showed no evidence of chromosome karyotypic aberrations when examined with comparative genomic hybridisation. SHIH *et al.* (2001) used single nucleotide polymorphisms to identify allelic imbalance in small colorectal adenomas. They identified frequent LOH on chromosomes 1, 5, 8,15 and 18, which they suggested was due to chromosomal instability, although the possibility that the pattern of LOH could simply be due to allelic loss, followed by reduplication, rather than gross genetic instability, cannot be discounted. Gross karyotypic abnormalities undoubtedly occur in a large number of colorectal cancers. It remains unclear though, whether these alterations are an indication of an underlying unstable genome, or the result of a one-off selectively advantageous event that, for example, removes a functional tumour suppressor gene or causes polyploidy.

1.7.2 Microsatellite instability

Microsatellites consist of a short, repeated, motif of bases and are found throughout the genome (DE LA CHAPELLE, 2003). For example, the sequence CACACACACACACACA

forms a dinucleotide microsatellite of length eight motif repeats.

Microsatellite instability (MSI) is a particular form of genetic instability that is relatively common in colorectal cancer. MSI is characterised by somatic alteration of the number of motif repeats at many microsatellite (MS) loci. The somatic alterations are due to *slippage*, whereby the DNA replication enzyme polymerase can "slip" a number of motif repeats when synthesising the repetitive microsatellite regions (DE LA CHAPELLE, 2003).

Whilst it is unlikely that the microsatellite instability has any selective effect upon a cell *per-se*, microsatellite slippage in a MS locus within a tumour suppressor or oncogene can cause a change in the protein structure, often a truncated protein product. In this manner, the acquisition of an MSI phenotype accelerates tumour progression. Indeed, patients with *Hereditary Non-Polyposis Colorectal Cancer* (HNPCC), who have inherited defects in their *mis-match repair* machinery, causing MSI, will tend to be diagnosed with a tumour around twenty years prior to the general population (KINZLER and VOGELSTEIN, 1996). HNPCC is discussed further below.

Detection of microsatellite slippage is via routine polymerase chain reaction (PCR) techniques. MS loci are first PCR amplified. Somatic alteration of the MS in the tumour, relative to the germline, can be identified by gel electrophoresis as changes, in the allele length affect the motility of the PCR product through the gel.

MSI-H

Around 15% of tumours have extensive MS slippage at many of the loci examined (DE LA CHAPELLE, 2003). Accordingly, these cancers are labelled *high-level microsatellite unstable* (*MSI-High*, or *MSI-H*). MSI-H cancers have genetic or epigenetic defects in their *mismatch repair* (*MMR*) machinery (DE LA CHAPELLE, 2004; HOEIJMAKERS, 2001), a mechanism that proof-reads and corrects errors in DNA replication (IYER *et al.*, 2006). Thus in MMR deficient cancers, MSI results from a persistent inability of the cancer cells to identify and repair replication errors (PARSONS *et al.*, 1993).

Patients with HNPCC, sometimes called *Lynch Syndrome*, have inherited defects in

MMR genes, and so exhibit MSI-H (KINZLER and VOGELSTEIN, 1996). The dominant germline mutations are typically in *MLH1* (PAPADOPOULOS *et al.*, 1994) and MSH2 (FISHEL *et al.*, 1993). Sporadic cancers identified as MSI-H often have somatic mutations in these genes, or, more often, show hyper-methylation of the hMLH1 promoter which is associated with protein loss (KANE *et al.*, 1997; VEIGL *et al.*, 1998).

MSI-L

In addition to the MSI-H cancers, it has been suggested that there is a second distinct group of cancers that are characterised by a low-level of microsatellite instability (termed *MSI-Low* cancers or *MSI-L*). Identification of MSI-L cancers is inherently problematic. Microsatellite loci are prone to slippage: estimates of the meiotic MS slippage rate per MS locus range from 10^{-9} to 10^{-6} (BHATTACHARYYA *et al.*, 1994; HANFORD *et al.*, 1998; KRUGLYAK *et al.*, 1998; LAI and SUN, 2003), compared to a baseline point mutation rate which may be as low as 10^{-10} per base per division (ELMORE *et al.*, 1983; SJÖBLOM *et al.*, 2006). Tumour growth involves many rounds of cell division, providing ample opportunity for MS slippages to accumulate in tumour cells. Thus, apparent low level microsatellite instability in a tumour may be attributable to normal somatic slippage that is expected during tumour growth. Accordingly, a large study, that genotyped 90 non MSI-H cancers at 377 loci, showed that if enough loci were examined, an unstable locus would nearly always be found (LAIHO *et al.*, 2002).

Nevertheless, many groups have attempted to characterise the group of putative MSI-L cancers. MSI-L status is usually identified by slippage by at least one locus, and usually less than 30% of the loci, examined (reviewed in TOMLINSON *et al.* (2002)). KETS *et al.* (2006) showed a weak association between MSI-L cancers and lymphocyte invasion. WHITEHALL *et al.* (2001) showed an association between methylation of *O 6-Methylguanine DNA Methyltransferase* a DNA repair enzyme, although a subsequent study failed to repeat this observation (HALFORD *et al.*, 2005). Suggestions that MSI-L was associated with K-ras mutation by JASS *et al.* (1999) again could not be verified in a subsequent study by LAIHO *et al.* (2002). MSI-L has not been shown to correlate

strongly with worse prognosis or any other clinico-pathological variable (WRIGHT *et al.*, 2005; HALFORD *et al.*, 2002).

MSI-L is considered in detail in Chapter 6. A mathematical model that predicts the degree of MSI expected in cancer from normal somatic slippage is developed, and used to assess the evidence for MSI-L in a cohort of colorectal cancers.

1.7.3 Clonal expansion versus genetic instability

Whilst genetic instability (GI) undoubtedly occurs in some colorectal cancers, such as in the tumours of HNPCC patients as noted above, it is hotly debated whether GI is necessary for tumorigenesis. Mathematical modelling has been used to attempt to answer this question. TOMLINSON *et al.* (1996) used a simple argument to suggest that GI was unlikely early event in tumorigenesis. They showed that if tumour growth required just two hits, and the acquisition of genetically unstable phenotype also required two hits, then cancers that formed rapidly were unlikely to have genetic instability, since this required the cancer to acquire four hits in a short period of time. If the number of mutations required for tumour growth was increased, then GI was more likely to be an early event in tumorigenesis.

A more sophisticated model was developed by Nowak and colleagues. Their stochastic model considered the APC mutation in a colon crypt (NOWAK *et al.*, 2002; KOMAROVA *et al.*, 2002). Wild-type cells could acquire a first APC hit at a low basal and the second hit occurred via LOH event or second point mutation. Alternatively, a wild-type cell could acquire mutations that caused chromosomal instability, which increased the rate of LOH in an $APC^{-/+}$ cell. In contrast to the previous work of Tomlinson and colleagues, Nowak's model also considered the dynamics of the stem cell niche in a colon crypt. Mutations had only a finite probability of fixation in the niche, and so could be lost from the niche. Using this framework, they showed that if the number of cells in a crypt susceptible to cancerous mutation was small, then CIN could precede APC inactivation. A caveat of this conclusion was that LOH occurs at a similarly low frequency to point mutations. If the rate of LOH was greater than the rate of point mutations, as may well be the case, then CIN was unlikely to be a initial event in tumorigenesis.

In their seminal work, ARMITAGE and DOLL (1954) attempted to estimate the number of mutations required to transform a somatic cell into a cancer cell. The modelling framework introduced by Armitage and Doll has subsequently been adapted to study the competing roles of mutation and clonal expansion in tumour growth. Armitage and Doll noted that a log-log plot of the number of cancer related mortalities against age followed an approximately straight line, for many cancer types. Thus the frequency of cancer related deaths was roughly proportional to the n^{th} power of age. For many cancer types, including colorectal cancer, the gradient of this line was about $n = 6$. Armitage and Doll showed that a multistage model of tumour progression, which required $n + 1$ changes to transform a normal cell to a cancer cell, was consistent with this data. In their model, the rate of acquiring a single mutation in time t was μt, and so the rate of at which a cell acquired $n + 1$ mutations was then asymptotically proportional to n^{th} power of age:

$$\text{formation rate} \sim \mu_1\mu_2\mu_3 \dots \mu_{n+1}t^n \tag{1.1}$$

Adapting the framework provided by Armitage and Doll, LUEBECK and MOOLGAVKAR (2002) described the effect of clonal expansion in a multistage model of tumorigenesis. After acquiring a pre-determined number of mutations, a mutant cell lineage could then initiate tumour growth. Tumour cells acquired additional mutations to become cancerous. The model was compared to age-incidence data on colorectal cancer formation. Fitting of the model suggested that the rate of the initial (APC) mutations was low - so genetic instability was unlikely to be an early event in tumorigenesis. All but one of the subsequent changes were also predicted to occur at a low rate, so that it was not necessary to invoke genetic instability to explain the cancer incidence data. In fact, models that use incidence data to infer the number of mutations required for malignant transformation depend strongly on underlying assumptions about clonal expansion. HORNSBY *et al.* (2008) demonstrated that including clonal expansion in a multistage model strongly affected the estimate of the number of alterations that occur during tumour growth. Furthermore, the clonal expansion considered in their model was minimal; a clone of size of around 150 cells was assumed to form after 50 years. This highlights the need for better understanding of the role of clonal expansion in tumour growth. In particular, understanding the role that clonal expansion plays in

determining the spectrum of mutations present in a cancer is important for determining the initial events in tumorigenesis. This is considered in Chapters 5 and 6 of this thesis.

Clear evidence for the importance of clonal expansion in driving tumour progression was demonstrated by BEERENWINKEL *et al.* (2007). Their model attempted to explain the presence of the large number of mutations observed in colon and breast cancers (SJÖBLOM *et al.*, 2006; JONES *et al.*, 2007). In their model, tumour growth began following APC inactivation. The tumour was assumed to grow at a constant rate, but the rate of subclone growth (or retardation) was determined by the relative fitness of the subclone, as specified by the subclone's genotype. A cancer was formed after at least one subclone had acquired $k = 20$ mutations. The authors found that it was not necessary to invoke an increased mutation rate in the model to explain the genetic diversity observed in cancers. Instead, it was sufficient to assume that each additional mutation in a subclone bestowed an additional selective advantage of around just 1%. Increasing the mutation rate accelerated cancer progression. Thus, selection, not mutation, drove cancer progression.

Whilst genetic instability may accelerate tumour progression, through increasing the rate of tumorigenic mutations, it is conceivable that it may also have a deleterious effect on the tumour cell population. Genome wide genetic instability presumably increases the rate of mutations in genes necessary for cell function, so called *housekeeping* genes. Thus a genetically unstable cell may be more likely to acquire mutations that make it less fit relative to the other cells in the tumour, and so in this manner, genetic instability may be selected against. Using a continuum model, where cell fitness was described as a continuous function of the mutation rate, Sole *et al.* predicted that a genetically unstable tumour population would be maintained by a few genetically stable *cancer stem cells*. However, the robustness of this conclusion is confounded by the fact that the authors did not permit mutation in the cancer stem cells, so it is unclear whether the stratification of the tumour population by genetic instability is a natural consequence of tumorigenesis, or an artefact of the model. The relationship between the mutation rate and the balance between positive and negative selection warrants further study.

Since mutations tend to occur during DNA replication and chromosome segregation, the large number of divisions involved in tumour growth can generate genetic diversity in a population of tumour cells. In this sense, the genetic diversity generated by clonal expansion may be analagous to the genetic diversity generated by a mutator phenotype. Chapter 6 of this thesis assesses the degree of genetic diversity that can be generated in a population of tumour cells, without an elevated mutation rate.

1.8 Conclusion

This chapter has described the basic biology of stem cells and of the colonic epithelium. The joint roles of mutation and selection in transforming normal somatic cells to invasive cancer cells have been discussed. The need to understand how mutations can arise and spread, both prior to and during tumour growth, has been highlighted. This need is addressed in the remainder of this thesis.

In Chapter 2, a model of the stem cell niche in the Drosophila testes is developed. This model is used to demonstrate how a sporadically mutated stem cell can be selectively advantageous in the niche.

In Chapter 3, clonal expansion in the human colon is considered. The rate of crypt fission, the probable primary mechanism of mutant clone expansion in the colon, is estimated. The importance of crypt fission in fixing mutations in single crypts is also explored.

In Chapter 4, the laboratory techniques used to analyse the clonal structure of small colorectal adenomas are described.

In Chapter 5, the clonal structure of a number of small colorectal adenomas is reported. The mutation status of the *APC* and K-ras genes was determined on a crypt-by-crypt basis in 12 sporadic adenomas. The implications of this clonal mapping for the order of, timing of, and relationship between APC and K-ras mutations is discussed.

In Chapter 6, a mathematical model of somatic microsatellite slippage during cancer growth is developed. The model is used to analyse the evidence for low-level microsatellite instability in a cohort of colorectal cancers.

Finally, in Chapter 7, the implications of this thesis for future research are discussed.

Chapter 2

Stem cell division in the Drosophilia Testis

2.1 Aims

- Examine stem cell division in the *Drosophila testis* as a case study for examining stem cell regulation and survival in human tissues.
- Predict the behaviour of sporadically occurring mutant stem cells in the Drosophila testes, which have acquired mutations in specific genes.
- Demonstrate that some mutations, that are insufficient to initiate tumour growth, can still rapidly reach fixation within a stem cell niche.

2.2 Chapter Summary

Stem cell regulation is critical for development and tissue homoeostasis. Further, most cancers are likely to originate from stem cells, since only stem cells have lineages that are long enough to be likely to acquire the multiple genetic mutations required for the initiation of neoplastic growth (CAIRNS, 2002; CALABRESE *et al.*, 2004). Stem cells have the self-renewing properties required for tumour growth, whereas more specialised cell types would require a more redical alteration of their phenotype to become oncogenic (PIERCE and SPEERS, 1988). Therefore, understanding the mechanisms responsible for stem cell regulation, and therefore stem cell survival, is

essential for understanding the initial events in tumour formation.

Stem cells are thought to be spatially defined by physical interactions with a local extra-cellular environment or *niche* (WATT and HOGAN, 2000). Multiple stem cells are present in a niche (BARKER *et al.*, 2007; YATABE *et al.*, 2001; GÖNCZY and DINARDO, 1996), and potentially one stem cell may acquire a selective advantage from a sporadic mutation that allows its progeny to become dominant in the niche. In this manner, oncogenic mutations, which by themselves are insufficient to initiate tumour growth, can accumulate in a stem cell population. Studying stem cell survival in humans is difficult; there are few reliable stem cell markers (MOLOFSKY *et al.*, 2003; LESSARD and SAUVAGEAU, 2003; BARKER *et al.*, 2007; SANGIORGI and CAPECCHI, 2008) and time course experiments are usually impractical.

In this Chapter, stem cell survival in the Drosophila testis was studied. The Drosophila testes is a relatively well characterised niche, and so provided an excellent model system to investigate stem cell survival. In particular, there were sufficient data to permit investigation of the behaviour of phenotypically altered stem cells within the niche. In the fly testis, stem cells are clustered around a *hub* of non-dividing cells (GÖNCZY and DINARDO, 1996). The hub secretes a niche-signal which maintains cell *stemness* (TULINA and MATUNIS, 2001; KIGER *et al.*, 2001). As cells move away from the hub, the strength of the niche signal decreases, causing cells to differentiate. Stem cell numbers are controlled by regulation of the angle of the mitotic spindle during cell division (YAMASHITA *et al.*, 2003, 2007). New daughter cells are produced away from the hub, so that they receive a reduced niche-signal and subsequently differentiate.

I have developed and analysed a mathematical model of stem cell survival in the Drosophila testes. The model was used to investigate the behaviour of phenotypically abnormal stem cells in an otherwise wild-type niche. The parameters of the model were estimated from published data (YAMASHITA *et al.*, 2007; WALLENFANG *et al.*, 2006) and verified against a third published data set (TULINA and MATUNIS, 2001). The model was then used to predict the selective advantage of sporadically mutated stem cell clones in the niche. Simulations of stem cells that had homozygous mutations in the *apc1* and *apc2* genes, which caused an impaired ability to correctly align their mitotic

spindle (YAMASHITA *et al.*, 2003, 2007), revealed that these mutations were selectively advantageous in a niche composed of otherwise wild-type cells. Further, the mutant cells were predicted to rapidly displace wild-type cells from the niche. Additionally, the simulations predicted that cells with mutations in the *cetrosonin* (*cnn*) gene, which also resulted in an impaired alignment of the mitotic machinery (YAMASHITA *et al.*, 2003, 2007), were also selectively advantageous.

Analysis of the model revealed properties of a stem cell that conferred a survival advantage in the Drosophila testes stem cell niche. The model demonstrated that stem cell mutations that may have oncogenic potential, but are insufficient to initiate tumour growth, can rapidly become fixed in a stem cell niche, and so predispose to further oncogenic development.

2.3 Introduction

2.3.1 Tissue structure and role of stem cells

Many tissues in the human body are rapidly renewed. For example, the cells that form the epithelium of the large intestine have a life-time of around one week until they are replaced (BRITTAN and WRIGHT, 2004b). Tissue renewal is driven by the division of *stem cells* (BRITTAN and WRIGHT, 2004a). Stem cells have the ability to divide either *symmetrically* or *asymmetrically*, producing, respectively, either two daughter stem cells or two differentiated cells, or a stem cell, and a *differentiated cell* (WATT and HOGAN, 2000). The differentiated cell can undergo further *amplifying divisions*, which produce the specialised cells that form the majority of a tissue or organ.

Understanding the behaviour of stem cells is therefore crucial to understanding development and homoeostasis. Dysregulation of stem cell division is also likely to be the initiating event in tumorigenesis, since only stem cells have a high replicative potential, and lineages that are long enough to be likely to acquire oncogenic mutations (PIERCE and SPEERS, 1988; CAIRNS, 2002). Knowledge of stem cell dynamics and crucially, stem cell survival within a niche, is therefore necessary to understand the first events in neoplastic growth.

Studying stem cell and niche interactions is difficult. There are few available stem cell

markers (reviewed in (WEISSMAN *et al.*, 2001)), and notably none which have been conclusively shown to identify stem cells in the human or murine colon (BARKER *et al.*, 2007; LEEDHAM *et al.*, 2005; SANGIORGI and CAPECCHI, 2008). Further, stem cells appear to be spatially defined by the location relative to the stem cell niche (WATT and HOGAN, 2000; WEISSMAN *et al.*, 2001). In human colonic crypts for example, stem cells are thought to reside near to the base of each crypt, proximal to the *pericryptal myofibroblasts*. The myofibroblasts are thought to secrete some members of the *Wnt* family of growth factors which reportedly promotes stemness (GREGORIEFF *et al.*, 2005). Cells differentiate as they migrate away from the crypt base, along the crypt axis. The inter-dependency of a stem cell and its niche means that, to study stem cell survival, observations must be made *in vivo* (THEISE, 2004), which makes it impractical to observe stem cell survival in humans.

A cell's intrinsic properties also affect its stemness. For example, germline loss of Tcf4, an integral component of the Wnt signalling pathway (see Introduction section 1.6.1), leads to the loss of all intestinal crypts in mice (KORINEK *et al.*, 1998), presumably because the cells can no longer respond to niche survival signals.

2.3.2 Stem cells in the Drosophila Testes

One well-studied stem cell compartment is in the Drosophila testis. In male fruit fly testis, about nine germline stem cells (GSCs) are clustered around a *hub* of non-dividing cells that express a *niche-signal*, the ligand *Unpaired* (*Upd*). Upd stimulates the *JAK-STAT* (Janus kinase-signal transducer and activator of transcription) signalling cascade that causes cells to retain their stem cell identity (KIGER *et al.*, 2001; TULINA and MATUNIS, 2001). Thus, proximity to the hub maintains stemness. As GSCs move away from the hub, the niche signal decreases and the cells differentiate (see Figure 2.1).

In the Drosophila testes, stem cell numbers are regulated by a simple mechanism that preferentially distributes daughter cells to specific locations in the niche. Positioning of daughter cells is controlled by the angle of the mitotic plane. YAMASHITA *et al.* (2003) showed that the stem cells usually align their mitotic spindle perpendicular to the hub surface, so that one daughter cell is produced close to the hub niche whereas

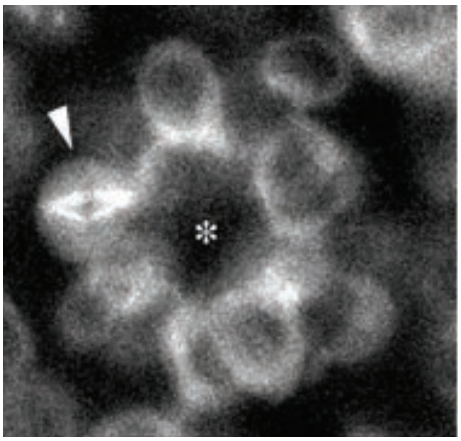


Figure 2.1: Micrograph of Drosophila testes hub and GSCs. The asterisk marks the position of the hub cells, and clustered around the hub are GSCs. The arrow indicates a dividing cell. The image is from a transgenic fly where α-tubulin content has been fluorescently labelled with green fluorescent protein (GFP). The mitotic spindle, with a high α-tubulin content is clearly visible in the dividing cell. Perpendicular alignment of the spindle to the surface of the hub is observed. Image reproduced with permission from YAMASHITA *et al.* (2003).

the other is produced at a distance and subsequently differentiates. In this manner, stem cell numbers remain constant within a niche.

Disruption to the mitotic machinery alters stem cell survival in the Drosophila testes. The testes of flies that were mutants for the integral centrosome component *centrosonin* (*cnn*) showed frequent misalignment of the mitotic spindle, causing both daughter stem cells to remain in contact with the hub after mitosis, with an overall increase in stem cell numbers (YAMASHITA *et al.*, 2003). Similar behaviour was observed in flies carrying mutations in *apc1* and *apc2* genes (YAMASHITA *et al.*, 2003). apc1 and apc2 are fly-homologues of the human APC gene (see Introduction 1.6.1), and are involved in anchoring microtubules which in turn orientate the centrosomes and the mitotic spindle (YAMASHITA *et al.*, 2003; ETIENNE-MANNEVILLE and HALL, 2003).

Another factor important for stem cell survival in the Drosophila testes is the ability of a stem cell to respond to niche signals. Disruption to components in the JAK-STAT signalling pathway that maintains GSC identity in the Drosophila testes causes loss of the stem cell population, whereas ectopic expression of the niche ligand Upd maintains

stem cell identity even when a cell is detached from the hub (KIGER *et al.*, 2001; TULINA and MATUNIS, 2001).

In this chapter, the importance of niche location and responsiveness to niche signals on stem cell survival in a model of the Drosophila testes stem cell niche is investigated. A mathematical model is developed which describes clonal evolution in the stem cell niche. The model is used to examine the ability of a mutant stem cell to completely displace wild type stem cells, and hence *invade* a niche. Using data on the behaviour of wild-type and mutant stems cell (TULINA and MATUNIS, 2001; YAMASHITA *et al.*, 2003, 2007; WALLENFANG *et al.*, 2006), predictions are made about how stem cells with specific types of mutation will behave in the niche.

2.4 Methods

2.4.1 Mathematical model of stem cell division

The model described N stem cells that were clustered around a central hub that secreted an attractive niche-signal. It was assumed that cells were adequately represented in two dimensions by circles of radius r. The strength of the niche signal was described by the function $w(h)$, where h was the distance of a cell from the hub. It was assumed that the strength of the niche signal decreased linearly with distance from the hub so that:

$$w(h) = 1 - \frac{h}{h_{max}} \tag{2.1}$$

where h_{max} was a normalising factor. At $h = 0$ a cell was just touching the surface of the hub and had a niche signal of 1.

At division, each stem cell produced a daughter cell at an angle θ to the radial line from the hub, so that the vertical distance of the daughter cell from the hub was $2r\cos(\theta)$ (see Figure 2.2). The niche signal for this daughter was then:

$$w(h) = 1 - \frac{2r\cos(\theta)}{h_{max}} = 1 - s\cos(\theta) \tag{2.2}$$

where s was a spatial scaling factor.

Each cell had a sensitivity α for the niche signal. A cell with a large value of α was influenced strongly by the niche signal. A cell with a large value of α was likely to be able to attach successfully to the hub.

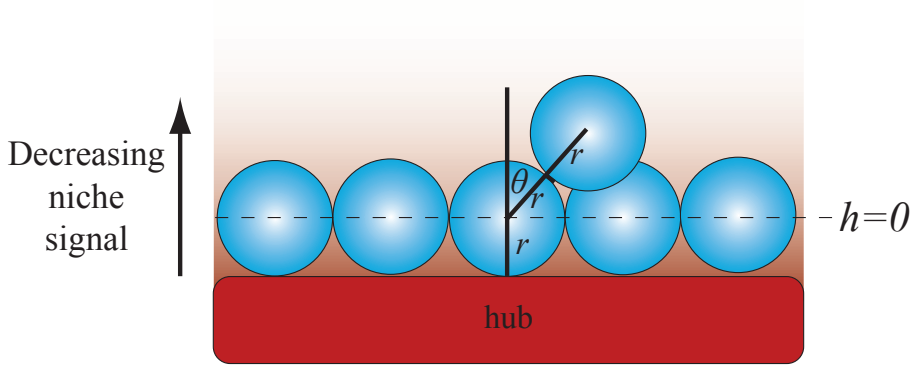


Figure 2.2: Cartoon of mathematical model construction. The hub of non-dividing cells secretes a *niche signal*, the ligand *Unpaired*. Stem cells in close proximity to the hub receive a strong niche signal, which stimulates the JAK-STAT pathway, and maintains stemness. Stem cells that move away from the hub receive a decreased niche signal and so differentiate.

The attraction A of a cell to the hub was then defined as the product of the sensitivity of the cell to the niche signal and the magnitude of the signal:

$$A = \alpha w(h) \tag{2.3}$$

Hence the attraction of the cell to the hub depended upon the distance of the cell from the hub.

After division, a daughter cell could occupy the position in the niche that was occupied by its nearest neighbouring stem cell. Competition for a place in the niche was decided by comparison of the attraction of the two competing cells for the hub. The cell that had the greatest attraction remained in the niche, whereas the other cell was lost from the niche and differentiated. After division, and the competition for spaces in the niche, precisely N cells remained in the niche.

The distribution of values of θ was modelled using a von-Mises distribution with mean $\theta_0 = 0$ and concentration parameter σ. Von-Mises distributions are the circular analogue of the normal distribution, periodic for $theta$ in the range $[0, 2\pi]$, so are appropriate for describing an angle. Then the probability density of θ was given by:

$$P(\theta|\theta_0 = 0, \sigma) = \frac{1}{2\pi I_0(\sigma)} e^{\sigma \cos\theta} \tag{2.4}$$

where I_0 was a modified Bessel function of the first kind. The concentration parameter σ controlled how peaked the distribution was. Large values of σ gave distribution that was strongly peaked around 0, whereas smaller values of theta gave a more uniform distribution. A cell with a larger value of σ therefore tended to align its mitotic machinery perpendicular to the surface of the hub and so produce a daughter cell at angle θ close to 0. Conversely, a cell with a small value of σ frequently misaligned its mitotic spindle.

2.4.2 Estimation of spindle alignment parameter σ_H

The mitotic spindle alignment tendency parameter for wild-type cells, σ_H, was estimated. To do this, it was assumed that each GSC would produce two GSC offspring at the same frequency that misaligned centrosomes were observed. The frequency of mis-orientated centrosomes observed in wild-type GSCs ranged between 1.3-12.2% (YAMASHITA *et al.*, 2003, 2007), but misaligned mitotic spindles were never observed in wild-type GSCs (YAMASHITA *et al.*, 2003). Therefore it was assumed that spindle alignment is difficult to score reliably and so underestimated the rate at which one GSC produced two GSC offspring. It was assumed that 6.1% of GSCs divisions produced two GSC offspring, according to the frequency of misorientated centrosomes in $\text{cnn}^{HK21}/+$ flies reported in YAMASHITA *et al.* (2007). Then, to estimate the corresponding value of σ_H, the model was repeatedly simulated, with increasing values of σ_H, to approximate the function $P(\text{correctly aligned centrosomes}) = f(\sigma_H)$ (see Figure 2.3). Values of σ_H were then calculated from this function. Additionally, without loss of generality, wild-type cell sensitivity was normalised to 1.

2.4.3 Estimation of stem cell division rate

Stem cell division rates were estimated by analysis of BrdU labelling data from WALLENFANG *et al.* (2006). Briefly, flies were fed on a yeast paste which had been mixed with Bromodeoxyuridine (BrdU). BrdU is incorporated into cells during DNA replication, and can be detected by immunohistochemistry. At t =12, 24, 36, and 48 hours of BrdU exposure, 10 flies were culled, and the mean proportion of BrdU labelled GSCs per testis was recorded. The proportions were labelled ρ_{12}^{Obs}, ρ_{24}^{Obs}, ρ_{36}^{Obs} and ρ_{48}^{Obs} respectively (see Table 2.1).

Hours of BrdU Exposure	Proportion of labelled stem cells
12	0.127
24	0.324
26	0.628
48	0.887

Table 2.1: Proportion of BrdU labelled GSCs against time. Data published in WALLENFANG *et al.* (2006).

A priori minimum (t_{min}) and maximum (t_{max}) durations of the cell-cycle were defined. t_{min} represents the minimum time required for cell division, whereas t_{max} represented an upper limit so that 99% of GSC cell divisions had a cell-cycle duration less than t_{max}. Estimates for t_{min} and t_{max} were 12 and 60 hours respectively.

The distribution of cell-cycle times for $T > t_{min}$ was then assumed to follow a log-normal distribution. In the absence of knowledge about the true distribution of cell-cycle times, the log-normal distribution was chosen because it has a Gaussian like shape, so that most cells had similar cell-cycle times, and also because the distribution was defined on the semi-infinite plane $[0, \infty)$ so that cell-cycle times were always at least as long as t_{min}.

The log-normal distribution has two parameters: μ and σ. Making the substitution $t' = t - t_{min}$, then the probability of a cell having a maximum cell-cycle duration of t was:

$$P(\text{cell-cycle duration T} \leq t) = P(T' \leq t') = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\ln(t') - \mu}{\sigma\sqrt{2}}\right) \tag{2.5}$$

Applying the *a priori* limits on the duration of the cell-cycle gave $P(T \leq t_{max}) = 0.99$. Hence equation 2.5 could be solved with $t = t_{max}$ to find:

$$\sigma = f(\mu) = \frac{\ln(48) - \mu}{\sqrt{2}\,\mathrm{erf}^{-1}(0.98)} \tag{2.6}$$

Hence the distribution of GSC cell-cycle durations (T) was defined as:

$$\begin{aligned} P(T < t) =& 0 && \text{where } t < t_{min} \\ P(T < t) =& \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\ln(t - t_{min}) - \mu}{\sqrt{2} f(\mu)}\right) && \text{where } t \geq t_{min} \end{aligned} \tag{2.7}$$

It remained to estimate the value of μ from the BrdU labelling data. When BrdU containing yeast paste was first made available to the flies, they did not immediately feed. In the absence of better knowledge, it was assumed that the duration between the availability of BrdU labelled yeast paste and the time when a fly began feeding was uniformly distributed on $[0, 24]$ hours. When flies were first exposed to BrdU, each GSC in a fly's testes could be at any point in its cell-cycle. Assuming that cell division was not synchronous, then the position in the cell-cycle of a GSC with cell-cycle duration T was described with a uniform distribution on $[0, T]$. The half-life of BrdU was neglected: the BrdU yeast paste was assumed to label cells with fixed efficacy.

A niche was initialised with $N = 9$ wild-type GSCs which had σ_H as estimated above. The time at which the fly ate the BrdU labelled food (A hours), was sampled from the uniform distribution on $[0, 24]$. For cell $i = 1, \ldots, 9$, the cell-cycle duration (T_i) was sampled from the distribution defined in equation 2.7 above. The position of cell i in the cell-cycle (C_i) was then sampled from a uniform distribution on $[0, T_i]$. Hence the division time of the cell was $D_i = T_i - C_i$. If $D_i \geq A$ then the *labelling time* of the cell was D_i, since the fly had consumed BrdU prior to the GSC division. Otherwise the labelling time was $D_i + \widehat{D_i}$, where $\widehat{D_i}$ was a second sample from the distribution of cell-cycle times in equation 2.7.

An Approximate Bayesian Computation method was then used to estimate μ (BEAUMONT *et al.*, 2002). For each repeat of the simulation, a value of μ was sampled from a uniform prior distribution on $[0, 10]$. This sampled value was labelled μ_S. Following the experimental method of Wallenfang *et al.* ten niches were then initialised as described above, using the sampled μ_S, and then simulated until $t = 48$ hours had elapsed. At t =12, 24, 36 and 48 hours the mean proportion of labelled cells in the ten niches was computed. These proportions were labelled ρ_1^{Comp}, ρ_2^{Comp}, ρ_3^{Comp} and ρ_4^{Comp} respectively. The value of μ_S was then accepted as an observation from the

posterior distribution of μ if

$$\delta \geq \sum_{j=1}^{4} \frac{|\rho_j^{Comp} - \rho_j^{Obs}|}{\rho_j^{Obs}} \tag{2.8}$$

where δ was a suitably small value. The simulation was repeated 10^6 times to form a posterior distribution for μ^{est} (see Figure 2.4). 95% credible intervals for μ^{est} were approximated by calculating the 2.5% and 97.5% percentiles of this distribution.

To evaluate the value of δ that provided an appropriately stringent acceptance threshold, the mean value and 95% credible points of μ were plotted as a function of δ. For $\delta < 0.2$ the mean and 95% credible points were approximately constant in value (see Figure 2.5), so did not depend strongly on the value of δ. Hence an *acceptance threshold* of $\delta = 0.2$ was used.

The estimated value of μ was $\mu^{est} = 3.23$ (2 dp) with 95% credible intervals 2.74 - 3.61 (2 dp). The fit of the model to the Wallenfang data is shown in Figure 2.6. This corresponded to a mean GSC division rate as one division every 39.4 hours (95% credible intervals 31.6 - 49.4 hours).

2.4.4 Verification of wild-type cell parameters

Using the parameters estimated above, the rate of loss of labelled wild-type cells, as reported in TULINA and MATUNIS (2001), was computed (displayed in Table 2.2). In TULINA and MATUNIS (2001), wild-type cells were labelled by heat shock at day 0. After 2 days, 55% of niches contained at least one labelled cell. It was assumed that the number of labelled cells per niche two days post labelling followed a Poisson distribution. Maximum likelihood estimation of the Poisson distribution parameter gave $\lambda = 0.7985$. Then, to compute the proportion of labelled niches after 2 days, the model was initialised at $t = 2$ days post-labelling, with L labelled cells in the niche, where L was sampled from the Poisson distribution estimated above. The model was then simulated and the fate of the labelled cells was recorded. The proportion of niches that contained a labelled cell after t was calculated by repeated simulation of the model ($R = 10^6$ repeats), and then counting the proportion of these R repeats that resulted in a niche containing a labelled cell at $t = 2, \ldots, 10$ days post-labelling.

Credible intervals for the proportion of labelled niches at t days post-labelling were

Days after induction	Proportion of testes with at least one labelled GSC
2	0.55
3	0.32
5	0.48
9	0.38

Table 2.2: Data from Tulina *et al.* on the proportion of labelled niches after heat shock induced labelling. Data reproduced from TULINA and MATUNIS (2001)

calculated using a bootstrap method. Following the experimental method of TULINA and MATUNIS (2001), at $t = 2$, 3, 5 and 9 days post-labelling, 10^5 sets of $n = 20$ niches were sampled, with replacement, from the $R = 10^6$ simulation repeats. The proportion of labelled niches in each set was then computed. These proportions approximated the sampling distribution of the number of niches containing a labelled cell at t days post labelling. 95% credible intervals were determined by calculating the 2.5% and 97.5% quantiles of the sampling distribution.

The fit of the model to the data from TULINA and MATUNIS (2001) is displayed in Figure 2.7.

2.4.5 Mutant Cells

Two types of mutant cells were possible in the model. Mutant cells could have a sensitivity to the niche signal, α_M, or a tendency to align their mitotic spindle σ_M, that was different to the wild-type stem cell population.

To examine the behaviour of mutant stem cells, a niche of host cells with parameters α_H and σ_H was constructed. At time $t = 0$ a single mutant cell with parameters α_M and σ_M was introduced into the niche by mutating one of the host cells. The lineage of this mutant cell was then tracked. For a given set of parameters the simulation was repeated many times ($R = 10^6$ repeats) and the proportion of runs where the mutant lineage successfully displaced all other cells in the niche was recorded. Also, the average

time to invasion and loss, and the *clonal stabilisation time* were computed. The clonal stabilisation time was the average time taken to either invasion or loss of the mutant cells.

Specific mutant genotypes

To predict the behaviour of a sporadic mutant stem cell with a specific mutation in either the cnn, apc1 or apc2 genes, the mitotic spindle alignment parameter σ_M for each mutant cell type was estimated. This was done according to the method for wild-type cells described above, using the centrosome misalignment frequencies reported in YAMASHITA *et al.* (2003). Estimated values of σ_M are shown in Tables 2.3 and 2.4, and are indicated on Figure 2.3. Additionally the simplifying assumption that mutant GSCs have the same sensitivity to the niche signal as wild-type cells was made. A niche that contained a single mutant cell at time $t = 0$ was then simulated, and the rate and likelihood of mutant cell invasion was examined. The likelihood and rate of mutant invasion for each mutant genotype are displayed in Tables 2.3 and 2.4.

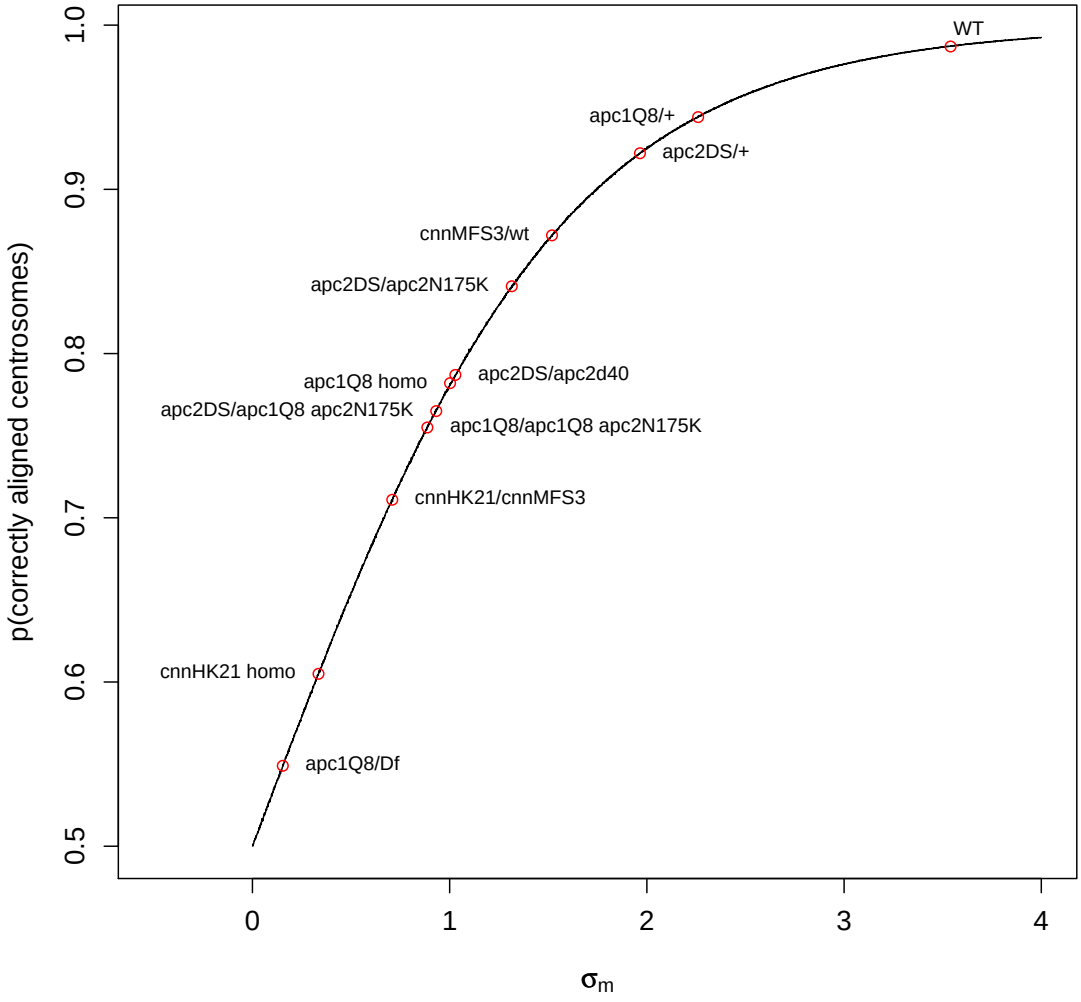


Figure 2.3: Proportion of GSC divisions with correctly aligned cetrosomes as a function of the mitotic spindle alignment parameter σ_H. The estimated values of σ_H for each strain of fly are indicated on the graph.

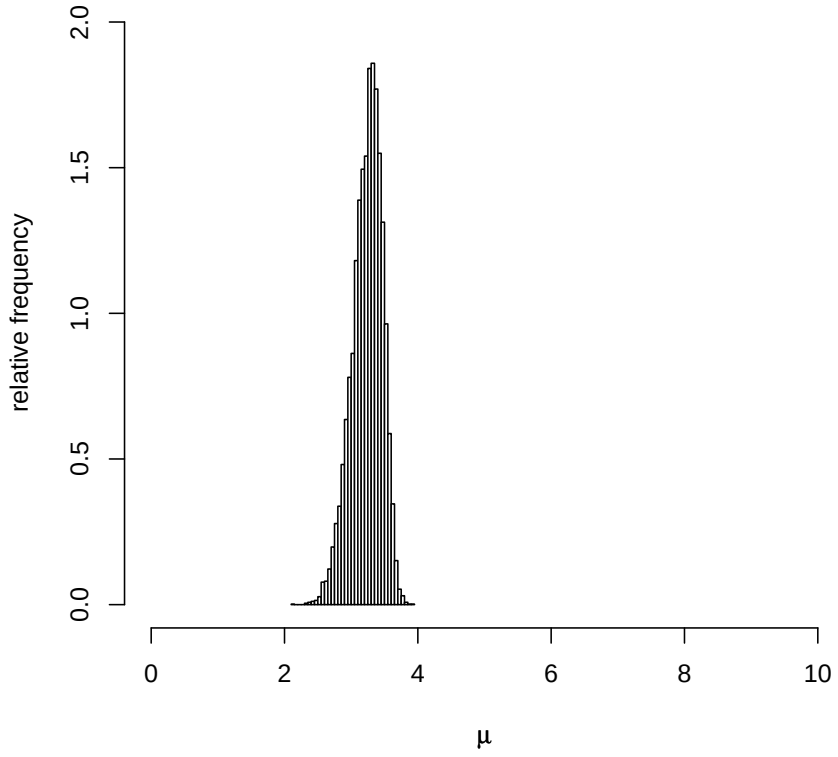


Figure 2.4: Posterior density for μ with an acceptance threshold of $\delta = 0.2$.

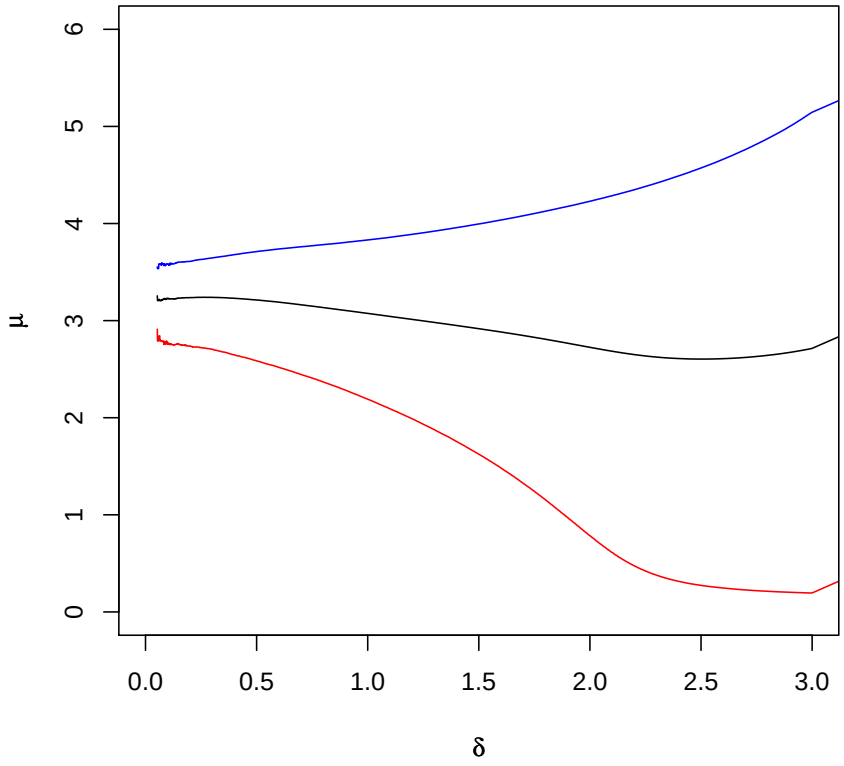


Figure 2.5: Estimated values of μ as a function of the acceptance threshold δ. The mean estimated value of μ is indicated by the black line, and 95^{th} percentiles are indicated by the red and blue lines. As the acceptance threshold was reduced $\delta \rightarrow 0$ the estimated values of μ converged to the true value under the posterior.

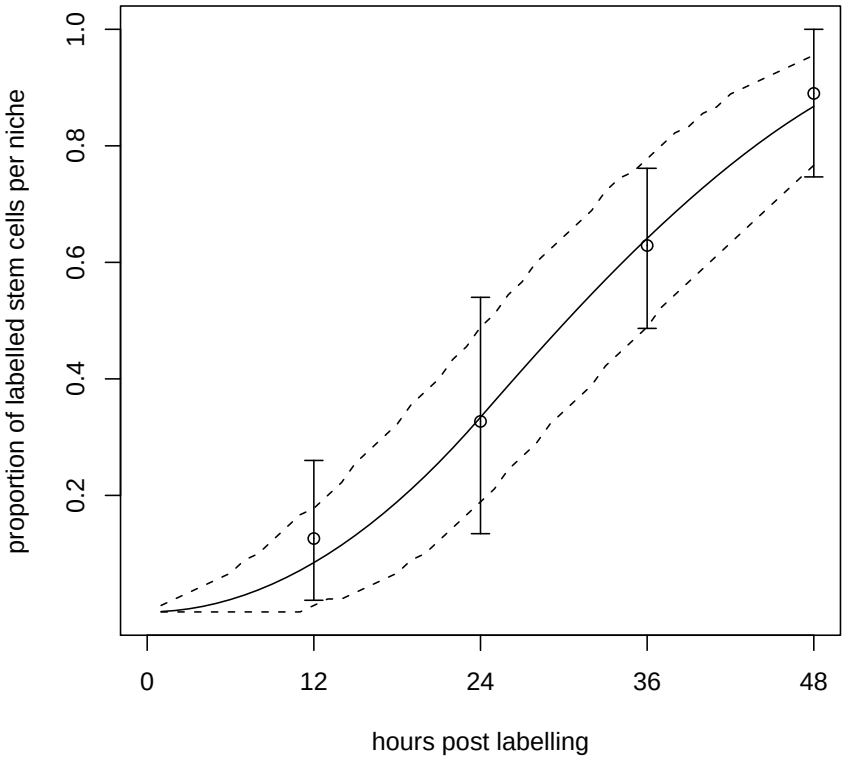


Figure 2.6: Fit of model to BrdU labelling data of Wallenfang *et al.* . Black circles show the proportion of labelled cells in a niche as measured by WALLENFANG *et al.* (2006). Solid lines are estimated number of labelled stem cells per niche with $\mu = 3.23$. Dashed black lines indicate 95% credible intervals for the fit.

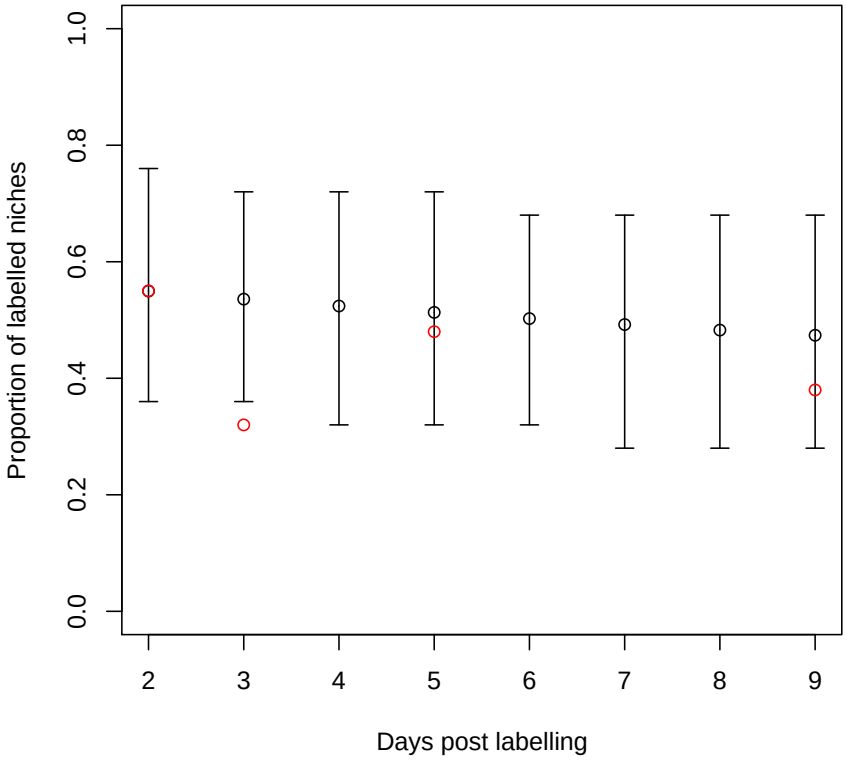


Figure 2.7: Fit of model to heat-shock induced labelling data of Tulina *et al.* . The figure shows the proportion of labelled testes against time, after induction of labelled cells at day 0. Red circles are data extracted from TULINA and MATUNIS (2001). Black circles are simulated values, with corresponding 95% credible intervals. The model predicts a steady decrease in the proportion of labelled niches over time, a trend generally supported by the data. However, the measured decrease in the proportion of labelled niche is more rapid than predicted; this is likely to be a result factors associated with an aging niche, that are not described in the model.

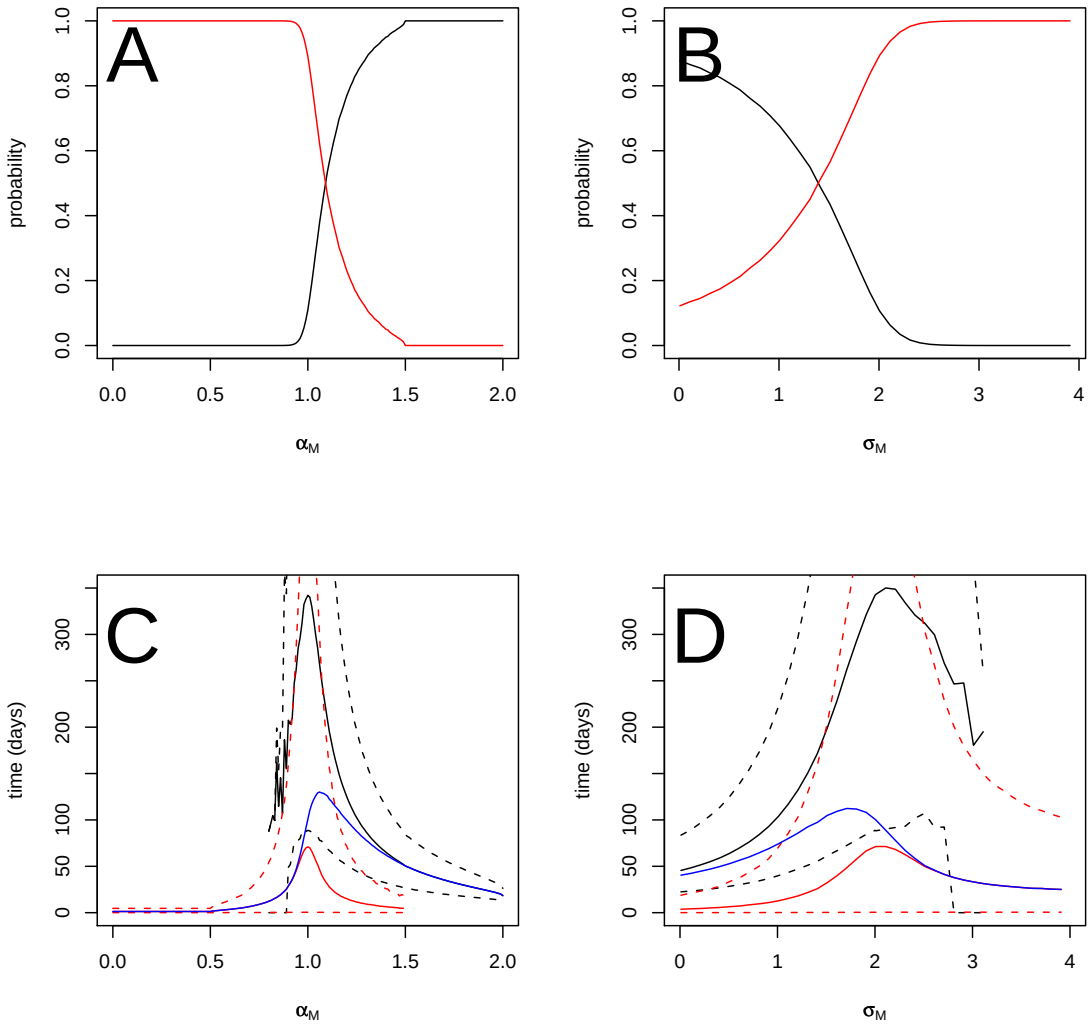


Figure 2.8: Behaviour of sporadic mutant cells in a niche. **A** and **B**: Probability of mutant invasion for, respectively, values of α_M, the sensitivity for the niche signal, and σ_M, the mitotic machinery alignment tendency. Red line, probability of mutant loss. Black line, probability of mutant invasion. Mutants were more likely to invade as α_M increased, and invasion was certain when α_M was sufficiently greater than α_H. Mutant invasion was also increasingly likely as the tendency for the mitotic machinery to misalign increased ($\sigma_M \to 0$). **C** and **D**: Times to mutant invasion and loss. Heavy red line: mean time to mutant loss. (95% credible intervals indicated by dashed red lines). Heavy black line: mean time for mutant invasion (95% credible intervals indicated by dashed black lines). Mutant invasion was a relatively slow process, with predicted invasion times often much longer than the average lifetime of a fly. Blue line: mean clonal stabilisation time. Parameter values: for all figures unless stated otherwise $\alpha_M = 1$, $\sigma_M = 2$, $\alpha_H = 1$, $\sigma_H = 2$, $N = 9$. Data displayed were calculated from 10^6 repeats of the simulation.

Table 2.3: Estimated invasion likelihood and rates of mutant cells (A).

Mutation	**P(asymmetric division)**	σ_m	**Mean Invasion Time (days)**	**Mean Loss Time (days)**
apc1^{Q8}/+	0.944	2.26	414.3 (108.2-1,107.1)	87.5 (0.6-586.5)
apc2^{DS}/+	0.922	1.965	336.8 (91.4-901.7)	66.2 (0.5-440.6)
cnn^{MFS3}/+	0.872	1.519	180.6 (60.4-421.3)	26.5 (0.3-157.0)
apc2^{DS}/apc^{2N175K}	0.841	1.315	138.3 (50.7-304.0)	18.1 (0.2-100.6)
apc2^{DS}/apc2^{d40}	0.787	1.029	98.6 (39.7-203.7)	11.6 (0.2-58.4)
apc1^{Q8}/apc1^{Q8}	0.782	1.002	95.7 (38.8-196.7)	10.9 (0.2-55.7)
apc2^{DS}/apc1^{Q8}apc2^{N175K}	0.765	0.932	88.9 (37.0-179.3)	9.9 (0.2-49.1)
apc1^{Q8}/apc1^{Q8}/apc2^{N175K}	0.755	0.887	85.0 (35.7-171.1)	9.3 (0.1-47.3)
cnn^{HK21}/cnn^{MFS3}	0.711	0.709	71.8 (31.6-139.7)	7.4 (0.1-36.1)
cnn^{HK21}/cnn^{HK21}	0.605	0.335	53.4 (25.5-99.0)	4.7 (0.1-22.5)
apc1^{Q8}/Df	0.549	0.154	47.5 (23.5-86.9)	3.9 (0.1-19.4)

Table 2.4: Estimated invasion likelihood and rates of mutant cells (B).

Mutation	**Clonal Stabilisation Time**	**40 days**	**Probability of eventual invasion**	**Probability of eventual loss**
apc1^{Q8}/+	112.3 (0.7-702.3)	0.000	0.076	0.924
apc2^{DS}/+	135.1 (0.6-674.1)	0.000	0.255	0.745
cnn^{MFS3}/+	109.6 (0.7-370.1)	0.001	0.539	0.461
apc2^{DS}/apc^{2N175K}	94.0 (0.7-278.4)	0.004	0.632	0.368
apc2^{DS}/apc2^{d40}	75.3 (0.7-193.3)	0.019	0.733	0.267
apc1^{Q8}/apc1^{Q8}	73.8 (0.6-186.7)	0.022	0.742	0.258
apc2^{DS}/apc1^{Q8}apc2^{N175K}	69.8 (0.7-171.5)	0.029	0.759	0.241
apc1^{Q8}/apc1^{Q8}/apc2^{N175K}	67.8 (0.7-163.9)	0.037	0.773	0.227
cnn^{HK21}/cnn^{MFS3}	59.7 (0.7-135.5)	0.074	0.811	0.189
cnn^{HK21}/cnn^{HK21}	46.9 (0.7-96.9)	0.223	0.867	0.133
apc1^{Q8}/Df	42.8 (0.7-85.3)	0.327	0.891	0.109

2.5 Results

2.5.1 Wild-type stem cell parameters

Stem cell division rate

The stem cell division rate γ and mitotic spindle alignment tendency parameter σ_H of wild-type cells was inferred. Each stem cell was estimated to divide once every 39.4 hours (95% credible intervals 31.6 - 49.4 hours) (see Figures 2.6 and 2.4).

Mitotic spindle alignment tendency

The alignment tendency of wild-type cells was estimated at $\sigma_H = 2.188$ (see Figure 2.3).

Fit to data of Tulina *et al.*

To confirm the estimation of these parameters, the model with these fuitted parameters, was compared to data from TULINA and MATUNIS (2001) on the loss of labelled wild-type cells from the niche (see section 2.4.4). Figure 2.7 shows the fit of the model, with parameters as stated above, to the data from TULINA and MATUNIS (2001). The model approximated the observed loss of wild-type cells well, although slightly over-estimated the number of labelled cells per niche two days post-labelling. This was likely to be because the number of GSCs per niche declines slowly with age (WALLENFANG *et al.*, 2006), but this phenomenon was not described in the model. Also, stem cells that had undergone recombination as a result of the heat shock treatment may have been less fit than other GSCs, so prone to loss from the niche. This was not considered in the model.

2.5.2 Invasion of mutant cells

The behaviour of sporadic mutant stem cells in the niche was investigated by exploring the behaviour of the model in α_M, σ_M parameter space. Mutant cells with a high sensitivity for the niche signal ($\alpha_M > \alpha_H$) were likely to invade, and correspondingly mutants with a low sensitivity for the niche ($\alpha_M < \alpha_H$) were unlikely to invade (see Figure 2.8A). Mutants invaded faster as their sensitivity to the niche signal increased

(see Figure 2.8C). Cells with disrupted mitotic machinery, which reduced the likelihood that they would successfully align their mitotic spindle perpendicular to the hub, were more likely to invade the niche than cells with their mitotic machinery intact. Invasion was most likely when there was no tendency for a mutant to align its mitotic spindle perpendicular to the niche ($\sigma_M \rightarrow 0$) (see Figure 2.8B). Mutants invaded faster as the mitotic spindle machinery became increasing disrupted (see Figure 2.8D).

2.5.3 Predicted invasion rate of stem cells with specific mutations

The behaviour of sporadic mutant GSCs with specific mutations in the apc1, apc2 or centrosonin (cnn) genes in an otherwise wild-type niche was investigated. For each mutant strain of stem cell, the tendency of the mitotic spindle to align correctly, σ_M, was estimated. The behaviour of a sporadic mutant stem cell in a niche was then simulated, and the likelihood and rate of mutant invasion were calculated (see Tables 2.3 and 2.4). cnn^{HK21}/cnn^{HK21} and apc^{1Q8}/Df mutants, both of which had severely disrupted mitotic machinery, were most likely to invade, with predicted mean invasion times of 53.4 (25.5-99.0) and 47.5 (23.5-86.9) days, respectively. Other mutant strains invaded more slowly, so that predicted invasion times were much longer than the average lifetime of a fly. Hence, these invasion times cannot be measured directly in a fly and therefore must be inferred.

2.6 Discussion

In this chapter, a mathematical model that describes the propagation of sporadic mutant stem cells in the Drosophila testes was developed. The model predicted that stem cells that had acquired mutations which disrupted the alignment of the mitotic machinery during cell division were likely to displace wild-type cells and hence become *fixed* in the niche. As the tendency of the mitotic spindle to align with the hub surface decreased, the likelihood that a mutated cell would invade increased. When the mitotic machinery was disrupted, a stem cell often produced a daughter cell close to the surface of the hub. The proximity to the hub provided the daughter cell with a strong attractive niche signal, so that the cell could compete favourably for a place in the niche.

Previously published work has shown that germline inactivation of apc1, apc2 or cnn

genes in the Drosophila testes results in stem cells with misaligned mitotic spindles (YAMASHITA *et al.*, 2003). The model developed in this chapter predicted that sporadic mutations of apc1, apc2 or cnn in a germline stem cell were selectively advantageous for the cell and increased the likelihood of the mutant cell invading the stem cell niche. The rates of invasion of cells with specific sporadic mutations were also predicted (see Tables 2.3 and 2.4). These predictions could be tested by experiment in a chimeric fly where apc1, apc2 or cnn could be selectively inhibited. The predictions were based on a niche which initially contained a single mutant stem cell, and the other cells in the niche were wild-type. In practice, induction of mutant cells in adult flies would tend to produce niches that initially contained more than one mutant cell per niche. It would be straightforward to modify the model to describe that situation.

The model also predicted that mutant stem cells that had acquired an increased sensitivity for the niche signal were likely to invade, as they were better able to compete with their neighbouring stem cells for a position in the niche. The *adenomatous polyposis coli* gene (*APC*), the mammalian homologue of apc1 and apc2, is implicated in cell migration and cell-cell adhesion (MAHMOUD *et al.*, 1997; FAUX *et al.*, 2004) so it is conceivable that mutations to apc1 and apc2 could confer increased sensitivity to the niche signal in fly testes. Correspondingly, in the testes of apc1 and apc2 mutant flies, increased numbers of stem cells are seen clustered around the hub (YAMASHITA *et al.*, 2003), which is indicative of an increased sensitivity to the niche signal, and correspondingly in testes where the niche signal ligand Unpaired was over-expressed stem cell numbers increased (TULINA and MATUNIS, 2001; KIGER *et al.*, 2001). Thus the actual selective advantage of apc1 and apc2 mutants (Tables 2.3 and 2.4) may be greater than predicted. Conversely, the model predicts cells with a low sensitivity to the niche signal are rapidly lost from the niche, which agrees with the observation that induced Stat92E mutant cells in adult flies, which cannot effectively transduce the Upd niche signal, are lost from the niche within a week (TULINA and MATUNIS, 2001; KIGER *et al.*, 2001).

Many mutations found in tumour cells may occur before visible tumorigenesis since some oncogenic mutations do not initiate tumour growth (KINZLER and VOGELSTEIN, 1997; JONES *et al.*, 2008b). This pre-tumour progression must occur predominately in

stem cells, since only stem cells have lineages long enough to be likely to accumulate the necessary multiple oncogenic mutations (CAIRNS, 2002; CALABRESE *et al.*, 2004). Stem cell niches contain multiple stem cells, and like any ecological niche that has an upper limit on the number of cells it can support, selection or drift is expected between similar but non-identical daughter cells. It is impossible to know *a priori* which stem cells will acquire oncogenic mutations before initiating tumour growth, therefore studying pre-tumour progression is prohibitively difficult. The model presented in this chapter illustrated how studying the behaviour of mutant cells in a simple model system can reveal stem cell phenotypes that promote stem cell survival. Stem cells in human niches may acquire mutations that are insufficient to initiate neoplastic growth, but which provide the stem cell with a selective advantage, which causes increased longevity of the mutant lineage. The model presented in this chapter illustrates how stem cells may acquire pro-oncogenic mutations which then become fixed in the niche. Therefore, the mutant stem cell lineage is more likely to acquire further mutations that may promote, and eventually prompt, cancer growth.

2.7 Conclusion

A simple mathematical model was developed that described the behaviour of sporadic mutant stem cells in the Drosophila testes. The model predicted that mutations which caused misalignment of the mitotic machinery of a stem cell were selectively advantageous in the niche, as were mutations which increased a cell's ability to respond to the niche signal and so compete favourably with other stem cells for a position in the niche. The model illustrated how mutations which may have oncogenic potential, but were insufficient to initiate tumour growth, could become rapidly fixed in a niche, and so predispose to subsequent progression to cancer.

2.8 Future work

2.8.1 Additional model complexity

The model developed in this chapter was designed to be a basal description of stem cell dynamics in the Drosophila testes, developed in order to investigate the behaviour

of sporadic mutant cells in a niche. In order to describe more precisely the true behaviours of germline stem cells in the Drosophila testes, additional behaviours could be introduced into the model. For example, the number of germline stem cells per niche decreases with age (WALLENFANG *et al.*, 2006), but this is not described in the model. Also, recent research suggests that the strength of the niche-signal is decreased in older flies (BOYLE *et al.*, 2007); this could also be described in the model. Further, mutations in apc1, apc2 and cnn genes alter the number of stem cells per niche. In the model presented in this chapter, the number of stem cells per niche is fixed. The model could be modified to allow fluctuations in stem cell numbers.

2.8.2 Experimental systems

In this chapter, the likelihood and rate of invasion of several sporadically occurring mutations have been predicted. The predictions could be tested in a suitable fly model, where specific mutations could be induced in adult flies. Specifically, flies with inducible mutations in apc1, apc2 or cnn should be created. Standard Flp-FRT is a suitable technology for induction of mutant clones, and could be used in a manner similar to the induction of STAT92E mutant clones described in (TULINA and MATUNIS, 2001).

Chapter 3

Crypt fission in the human colon

3.1 Aims

- Describe the spread of mutant clones in the human colon
- Estimate the duration of the crypt cycle.
- Examine the effect of crypt fission on the expansion of mutant clones.

3.2 Chapter Summary

Cancer can be considered a stem cell disease, since it is argued that only stem cells have lineages long enough to be likely to accrue sufficient mutations to initiate tumorigenesis. Understanding how mutations can arise and spread in a stem cell population is therefore crucial to understanding the early events in tumour formation. In this chapter, the spread of mutations in the stem cell population of the human colonic epithelium is examined.

The lining of the colon is covered in small invaginations called crypts. Each crypt is maintained by a number of stem cells that are thought to reside close to the crypt base, in the SC niche (POTTEN and LOEFFLER, 1990; SPRADLING *et al.*, 2001).

Recently, the mechanism of clonal expansion of colonic stem cells has been elucidated (GREAVES *et al.*, 2006). Using mitochondrial DNA mutations as clonal markers, it was demonstrated that a mutant stem cell can displace the other, wild-type, stem cells from the niche, and so become dominant in the niche. This clonal expansion process is

termed *niche succession* (YATABE *et al.*, 2001; CAMPBELL *et al.*, 1996). Subsequent clonal expansion is through *crypt fission*, whereby one parent crypt divides to produce two daughter crypts. The period of time between fission events defines the crypt life-cycle, termed the *crypt cycle* (TOTAFURNO *et al.*, 1987). The duration of the crypt cycle is of considerable interest, particularly because crypt fission is the primary mode of expansion of colorectal adenomas (PRESTON *et al.*, 2003; BJERKNES *et al.*, 1997). The previous estimate of the duration of the crypt cycle in humans was based on a heuristic comparison between murine and human crypts, and suggested that the number of crypts in the colon doubles every 9-18 years (CHENG *et al.*, 1986a).

In this Chapter, the duration of the crypt cycle is inferred from data on the spread of cytochrome-c-oxidase (COX) deficient crypts through the colonic epithelium ((GREAVES *et al.*, 2006) and unpublished data) using a parsimonious ordinary differential equation (ODE) model of the clonal expansion in the colon. Some mtDNA mutations cause a cell to become COX-deficient. COX-deficiency is readily detected using a histochemical stain (TAYLOR *et al.*, 2003). A COX-deficient cell can repopulate an entire crypt through niche succession. Observations show that COX-deficient crypts form at a low rate throughout adult life (TAYLOR *et al.*, 2003). Since the rate at which entire crypts become COX-deficient is low, spatially adjacent COX-deficient crypts are likely to share a common ancestor. This was confirmed by direct sequencing of the entire mitochondrial genome: adjacent COX-deficient crypts shared identical mtDNA mutations that were not seen in the surrounding wild-type crypts (GREAVES *et al.*, 2006).

Fitting the model to the data suggested that the mean duration of the crypt cycle is 23.6 years (CI: 17.3-32.6 years). This estimate is longer than the previous available estimate of 13-26 years (calculated from an estimated *crypt doubling time* of 9-18 years) (CHENG *et al.*, 1986a). Further, the model suggests that crypt fission increases the likelihood that a mutant cell will become fixed in the epithelium. Therefore, fission causes oncogenic mutations, which themselves may be insufficient to initiate tumour growth, to become fixed in a crypt. This predisposes to subsequent adenomatous growth. The low rate of fission in wild-type crypts may be an anti-tumorigenic mechanism designed to restrict the spread of mutant clones in the colon. The increase

in the fission rate in adenomatous growths is therefore likely to be a pivotal event in colorectal tumorigenesis.

3.3 Introduction

The colon, the last section of the gastrointestinal tract, has a rapidly renewing epithelium. Estimates suggest that cells on the epithelium have a lifetime around one week in the mouse large bowel (POTTEN and LOEFFLER, 1990). The surface of the colon is lined with millions of small invaginations called *crypts* (see Introduction, Section 1.3). Each crypt contains stem cells which are thought to be located towards the crypt base (BRITTAN and WRIGHT, 2004b). Stem cells can undergo asymmetric division to produce a daughter stem cell and a daughter differentiated cell. Subsequently amplifying divisions of the differentiated cells maintain the crypt population (BJERKNES and CHENG, 1999). No definitive marker for colonic stem cells has, as yet, been identified, and various indirect assays have been used to estimate the number of stem cells per crypt; estimates range from 4 to around 30 (see Introduction, Section 1.3.4).

Within a crypt, stem cell division must be tightly regulated to ensure homoeostasis, and like any constrained population in an ecological niche, stem cells undergo random drift, whereby one stem cell lineage will replace all the others (YATABE *et al.*, 2001). This replacement process is termed *niche succession*. A neutral mutation arising in the stem cell pool, which results in a *partially mutated* crypt, can reach fixation in the crypt via this process, and so produce a *wholly mutated* crypt.

Crypts are dynamic structures. Crypts can divide through a process termed *crypt fission*, which is analogous to cell division. Fission begins at the base of the crypt, splitting the stem cell pool between the two daughter crypts, and proceeds along the vertical crypt axis (see Figure 3.1). Presumably, subsequent rounds of stem cell divisions restore the normal complement of cells in each daughter crypt.

Colonic cancers are thought to originate in colonic crypts (see reviews (POTTEN and LOEFFLER, 1990; PRETLOW and PRETLOW, 2005)). Adenomatous growth follows the inactivation of the APC tumour suppressor gene (KINZLER and VOGELSTEIN, 1996;

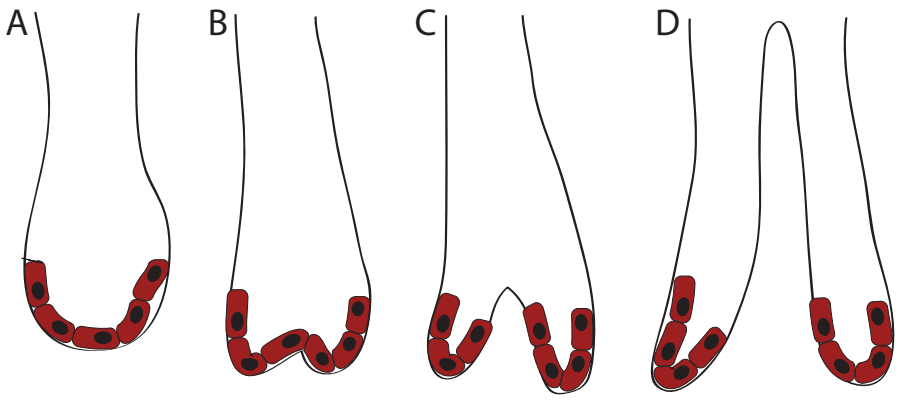


Figure 3.1: Cartoon showing process of crypt fission. Red cells represent stem cells. A-B: Fission begins at the base of the crypt and proceeds vertically along the crypt axis. C: Stem cells are divided between the two daughter crypts. D: Subsequent rounds of cell division restore the stem cell complement of each daughter crypt. Two complete crypts are formed.

LAMLUM *et al.*, 2000). The mutant $APC^{-/-}$ cell lineage is thought to repopulate the entire crypt. Two theories have been suggested for the subsequent expansion of the mutant crypt, *top-down* expansion, whereby cells from a mutant crypt invade a neighbouring crypt by downward migration from the crypt lumen (SHIH *et al.*, 2001), and *bottom-up* expansion (PRESTON *et al.*, 2003) whereby mutant clones form at the crypt base, and subsequent expansion of the mutated clone is by crypt fission. Both mechanisms of expansion do occur (see examples in Chapter 5 of this thesis). However, it appears that the initial expansion of colorectal adenomas is predominately through crypt fission, and later growth also involves top-down spread (PRESTON *et al.*, 2003). The number of crypts in fission is increased around sixteen-fold in adenomas (PRESTON *et al.*, 2003; WONG *et al.*, 2002), which is indicative that an increase in the rate of crypt fission in a mutated crypt is an important early development in tumorigenesis in the colon. Thus the regulation of crypt fission in a healthy bowel is likely to be a crucial homeostatic control.

The method responsible for the initiation of fission is unknown. TOTAFURNO *et al.* (1987) noted that crypts in the murine colon had a Gaussian-like distribution of sizes, and that crypts in fission tended to be larger than average. Subsequent observations by WONG *et al.* (2002) confirmed this. This has led some authors to speculate that

crypt fission occurs in response to changes in the stem cell number in the crypt. A mathematical model where fission occurs if the number of stem cells per crypt exceeded a threshold value (S_f) could explain the formation of adjacent mutant crypts in response to chemical induced damage in the mouse bowel (LOEFFLER *et al.*, 1993, 1997). Biomechanical modelling suggests that the increased cell number could causes *buckling* of the epithelial cell layer of the crypt relative to the underlying lamina propria, which could be the trigger to initiate fission (EDWARDS and CHAPMAN, 2007).

The distribution of crypt sizes, and the fact that around 3% of crypts in the murine colon are branching at any instant (TOTAFURNO *et al.*, 1987), led Bjerknes and Cheng, in their series of seminal papers, to suggest that crypts had a natural life-cycle, termed the *crypt cycle* (CHENG *et al.*, 1986b,c; BJERKNES, 1986; TOTAFURNO *et al.*, 1987). The crypt cycle defines the period of time between successive fission events in a single crypt. To estimate the duration of the crypt cycle in humans, CHENG *et al.* (1986b) assumed that the number of crypts in the colon was growing exponentially such that:

$$N(t) = A_0 e^{\alpha t} \tag{3.1}$$

(where $N(t)$ was the number of crypts in the colon at time t, α was the growth rate of crypts, and A_0 the number of crypts at $t = 0$). Setting $\alpha = \log(2)/C$ they had that the number of crypts doubles in time C. The time C was defined as the mean duration of the crypt cycle. Rearranging equation 3.1 for C gave:

$$C = \frac{t \log(2)}{\log\left(\frac{N(t)}{N(0)}\right)} \tag{3.2}$$

Then to estimate the crypt doubling time, it remained to estimate the change in the number of crypts in the interval $[0, t]$. Examination of the colons of 11 patients suggested that the percentage of crypts in the human colon that were in fission at any one time was 0.44%. Observations of the murine colon suggested that duration of a fission event in the mouse was between 5 and 10 days, and that the crypt stem cell cycle duration was between 15-23 hours. Cheng and Bjerknes estimated the frequency of stem cell division was four-fold greater in the murine colon than in the human, and so assumed that the duration of crypt cycle would also be four-fold greater in the human. Therefore, in a period of between 20 and 40 days, the proportion of crypts in the human

colon would increase by 0.44%. Inserting these numbers into equation 3.2 yielded an estimate of the duration of the *crypt doubling time* as between 9-18 years. This heuristic estimate has remained the *de facto* estimate of the crypt cycle duration in the human colon.

Crypt extinction or death is not thought to occur. Observations over a three week period of the same section of murine colon in three mice showed no evidence of crypt loss (BJERKNES, 1986), although the reliability of these experiments is questionable. It is difficult to confirm reliably the identity of individual crypts in the malleable bowel, and also extinguished crypts may have been rapidly replaced so that dying crypts may not have been observed. Given the large number of crypts in the colon, around 10^7, it seems very unlikely that occasional stochastic crypt death does not occur, especially since inflammation is associated with increased levels of oxidative stress (reviewed in PAVLICK *et al.* (2002)) which can cause cell apoptosis (reviewed in KANNAN and JAIN (2000)).

In this chapter, the duration of the crypt cycle is inferred from the data on the spread of mutated clones through the epithelium. *Cytochrome-c-oxidase* (COX) deficiency, which usually results from a mutation in the mitochondrial genome, was used as a lineage marker to trace the spread of mutated clones. COX-deficiency is not thought to confer a selective advantage on a cell (MCDONALD *et al.*, 2006b).

COX deficiency is readily detected using two-colour histochemical staining (TAYLOR *et al.*, 2003). After staining, COX-deficient cells appear blue, and wild-type cells appear brown. Greaves *et al.* used COX-deficiency to trace the expansion of mutated clones through the colonic epithelium. COX-deficient cells could form in a crypt, producing a partially mutated crypt (PMC), and then repopulate the entire crypt, forming a wholly mutated crypt (WMC). A WMC could then undergo fission, producing a *patch* of WMCs. Patches expanded through subsequent fission events. The mean patch size increased with age, which was taken as evidence of the crypt-cycle.

COX mutations occurred relatively infrequently, and were very rarely in seen patients aged less than 40 years. COX-deficient crypts that were spatially adjacent were likely to

share a recent common ancestor. This was confirmed by direct sequencing of the entire mitchondrial genome in single cells dissected from adjacent mutated crypts. Adjacent crypts shared the exact same mutation; the probability of this happening at random was estimated to be 2.48×10^{-9} (GREAVES *et al.*, 2006). A micrograph of a patch of COX-deficient crypts is shown in Figure 3.2.

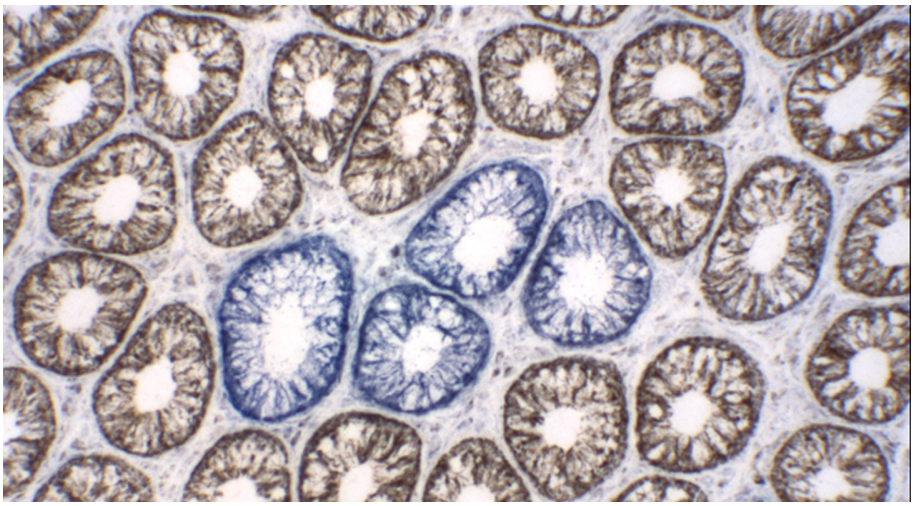


Figure 3.2: A patch of COX-deficient crypts in a section of human colonic mucosa. COX-deficiency is revealed as a patch of blue crypts, surrounding wild-type crypts are stained brown. Picture supplied by Dr Stuart McDonald, Cancer Research UK London Research Institute.

In this chapter, data on the spread of COX-deficient clones through the human colon was used to estimate the duration of the crypt cycle. In contrast to the previous estimate, the inference of the crypt fission rate in this Chapter is basely solely of data from the human bowel. A parsimonious mathematical model of mutation, fixation and crypt fission was developed. The model predicted the incidence of COX-deficient crypts with patient age. Fitting the model to the data yielded an estimate of the crypt fission rate. This estimate was confirmed using a more simplified model which produced a similar estimate. Subsequent analysis was performed to assess the importance of crypt fission in fixing mutations in the colonic epithelium.

3.4 Method

3.4.1 COX-deficiency data

Data were collected from colon specimens from 62 patients aged between 25 and 85 years. Data on 48 of these patients had been previously published (GREAVES *et al.*, 2006), and data on the remaining 14 patients were collected by Stuart McDonald (CRUK, Histopathology Laboratory, Lincoln's Inn Fields). COX deficiency was identified using an histochemical method as described in GREAVES *et al.* (2006) and TAYLOR *et al.* (2003). Briefly, colonic sections were cut to a thickness of 4μm and slides were dewaxed as described in Section 4.3.1. Antigen retrieval was performed by pressure cooking the slides for 3 min in boiling 0.1 M sodium citrate buffer (pH 6.0). Non-specific staining was blocked by adding protein block serum-free ready-to-use buffer (DAKO) for 15 min, then slides were washed three times in Tris-buffered saline with 0.1% Triton X-100. A 1:100 dilution of mouse anti-human cytochrome c oxidase subunit I was made in Tris/Triton X-100 and applied to the sections for 30 min in a humid chamber. Next, slides were washed three times in TBS/Triton, and then incubated in neat CheMate (Envision) anti-rabbit/anti-mouse conjugated with horseradish peroxidase (DAKO) and developed in 4 mM diaminobenzidene tetrachloride (DAKO). The sections were lightly counterstained in haematoxylin, dehydrated through increasing ethanol washes, cleared with xylene, and mounted in DPX.

In each specimen, the total number of crypts, and the frequency of wholly-mutated crypts (COX-deficient in a part of the crypt), and patches of mutated crypts (wholly COX-deficient), classified according to the size of the patch, was recorded. This data are displayed in the Appendix, Section 8.1.

Counts of partially mutated crypts were also occasionally recorded. These counts are considered unreliable and were neglected from this analysis for the following reasons. Firstly, distinguishing between a partially mutated crypt and a crypt which contained cells with a varying level of heteroplasmy for COX-deficiency was not possible. Secondly, confident identification of a partially mutated crypt was not necessarily evidence that the crypt contained COX-deficient stem cells, as COX-deficiency may

have been acquired by a long-lived commited progenitor cell which would eventually be lost from the crypt (see examples in HUMPHRIES and WRIGHT (2008)). Thus, including these partially mutated crypts would confound the analysis of crypt stem cell dynamics.

3.4.2 Summary statistics

To describe those data, two summary statistics were used: the proportion of counted crypts that were COX-deficient ($\phi_1(t)$), and the mean number of mutated crypts in a patch ($\phi_2(t)$). For patient j, aged t_j, there were τ_j crypts counted, M_j of which were COX-deficient. The total number of patches in the sample was denoted by P_j. Therefore:

$$\begin{aligned} \phi_1(t_j) &= \frac{M_j(t_j)}{\tau_j(t_j)} \\ \phi_2(t_j) &= \frac{M_j(t_j)}{P_j(t_j)} \end{aligned} \tag{3.3}$$

3.4.3 Deterministic model to estimate crypt fission rate

An ordinary differential equation (ODE) model (Model 1) was used to describe the propagation of mutations in the colon. Crypts containing a COX-deficient cell were assumed to form from wild-type crypts at a constant rate of λ_m per crypt per year, in colons aged more than T years. These *partially mutated* crypts became *wholly mutated* at a rate of λ_c per crypt per year. All crypts divided at a constant rate of λ_f per crypt per year: thus it was assumed that a crypt becoming COX-deficienct did not alter its fission rate.

It was assumed that all partially mutated crypts would become wholly mutated; thus the parameter λ_m represented the formation rate of COX-deficient crypts that were destined to become wholly mutated. The true rate at which COX-deficient cells formed in a crypt would have been higher than λ_m, since some COX-deficient cells would not have successfully invaded the stem cell niche, and therefore would have been eventually lost from the crypt. The assumption that COX-deficient cells formed at a constant rate from wild-type crypts was almost certainly an over-simplification (see Section 3.3), but was used as an adequate first approximation in the absence of information on the mutation and expansion rate of mtDNA strands containing COX mutations.

The dynamics of the stem cells in the niche were described by the single rate parameter λ_c. More sophisticated models of niche dynamics have been proposed (see for example YATABE *et al.* (2001); LOEFFLER *et al.* (1997, 1993); MEINEKE *et al.* (2001)), but the basal model here was assumed to be sufficient to reasonably describe the conversion from a partially mutated to a wholly mutated crypt. The dependence of the results on the value of λ_c was examined.

Under the assumptions of the model, the number of crypts in the colon grew exponentially at a rate of λ_f per crypt per year. Previous modelling attempts have speculated that crypt fission occurs in response to increased stem cell numbers in the crypt (TOTAFURNO *et al.*, 1987; LOEFFLER *et al.*, 1997; EDWARDS and CHAPMAN, 2007). In this model, the process that initiates fission was neglected, and it was assumed that fission events occur at a constant rate. This basal model allows the straightforward estimation of the crypt cycle duration, without requiring the postulation of the mechanism responsible for initiating fission.

Since relatively few partially mutated crypts are observed, the value of λ_c is likely to be much greater than λ_f or λ_m. Accordingly, a simpler model (Model 2) was developed which neglected the process of monoclonal conversion. In this second model, wholly mutated crypts were formed at a rate λ_m per crypt per year from wild-type crypts. All crypts underwent fission at a rate of λ_f per crypt per year. This second, simpler, model was used to evaluate the results of the first model and indicate the independence of the fitted value of λ_f to the monoclonal conversion rate λ_c.

Table 3.1: Definition of variables and parameters

Variable	Definition	Units
$N(t)$	number of wildtype crypts	
$Q(t)$	number of partially mutated crypts	
$M(t)$	number of wholly mutated crypts	
$P(t)$	number of patches of mutated crypts	
$\tau(t)$	total number of crypts	
λ_m	rate of formation of partially mutated crypts	per crypt per year
λ_c	monoclonal conversion rate	per crypt per year
λ_f	rate of crypt fission	per crypt per year
T	time of formation of first wholly mutated crypts	years
t	transformed time	years
$\phi_1(t)$	proportion of mutated crypts	
$\phi_2(t)$	mean patch size	

Parameter Definitions

3.4.4 Model 1: Definition

The model described above was defined by the following equations:

$$\frac{dN}{dt} = \lambda_f N - \lambda_m N \tag{3.4}$$

$$\frac{dQ}{dt} = \lambda_m N + \lambda_f Q - \lambda_c Q \tag{3.5}$$

$$\frac{dM}{dt} = \lambda_c Q + \lambda_f M \tag{3.6}$$

$$\frac{dP}{dt} = \lambda_c Q \tag{3.7}$$

$$\frac{d\tau}{dt} = \lambda_f \tau \tag{3.8}$$

with the initial conditions:

$$\tau(0) = N(0) = N_0; Q(0) = M(0) = P(0) = 0 \tag{3.9}$$

Time $t = 0$ was a transformed time, such that $t = t' - T$, where t' was the true age, in years, of the colon. Equations 3.4 - 3.8 describe the average behavior of the model.

The growth rate of each population of crypts (wild-type, partially mutated, or wholly mutated) is λ_f per crypt per year. Accordingly, the total number of crypts in the colon, $\tau(t)$, increases at rate λ_f per crypt per year (Equation 3.8). Wild-type crypts, $N(t)$, become partially mutated at rate λ_m per crypt per year (Equation 3.4). Monoclonal conversion of partially mutated crypts, $Q(t)$, occurs at rate λ_c per crypt per year - these crypts are then wholly mutated, $M(t)$, (Equations 3.5 and 3.6). Since each partially mutated crypt that is converted to wholly mutated is assumed to form a distinct (new) patch of COX-deficient crypts, the increase in the number of patches over time is simply the rate at which partially mutated crypts convert to wholly mutated (Equation 3.7). The initial conditions state that at time $t = 0$ there are N_0 crypts in the colon, and all crypts are wild-type and so there are no mutated crypts or patches of mutated crypts.

The number of wholly mutated crypts was calculated as the total number of crypts minus the number of wild-type and partially mutated crypts:

$$M(t) = \tau(t) - N(t) - Q(t) \tag{3.10}$$

Expressions for $\phi_1(t)$ and $\phi_2(t)$

Integrating (3.4) gives:

$$N(t) = N_0 \exp\left[(\lambda_f - \lambda_m)t\right] \tag{3.11}$$

Similarly, integrating (3.8) gave:

$$\tau(t) = N_0 \exp\left[\lambda_f t\right] \tag{3.12}$$

Substituting $N(t)$ into equation (3.5) and integrating gave:

$$Q(t) = N_0 \frac{\lambda_m}{\lambda_c - \lambda_m} \left(\exp\left[(\lambda_f - \lambda_m)t\right] - \exp\left[(\lambda_f - \lambda_c)t\right] \right) \tag{3.13}$$

Using the expression for $M(t)$ given in equation (3.10):

$$M(t) = N_0 \exp[\lambda_f t] \left(1 - \exp[-\lambda_m t] - \frac{\lambda_m}{\lambda_c - \lambda_m} \exp[-\lambda_m t] + \frac{\lambda_m}{\lambda_c - \lambda_m} \exp[-\lambda_c t] \right) \tag{3.14}$$

Then, dividing by the expression for $\tau(t)$ gave the expression for the proportion of wholly mutated crypts:

$$\phi_1(t) = 1 - \exp[-\lambda_m t] - \frac{\lambda_m}{\lambda_c - \lambda_m} \exp[-\lambda_m t] + \frac{\lambda_m}{\lambda_c - \lambda_m} \exp[-\lambda_c t] \tag{3.15}$$

Calculation of the mean patch size, $\phi_2(t) = M(t)/P(t)$, required the integration of equation (3.7). Substituting the expression for $Q(t)$ into (3.7) and integrating gave:

$$P(t) = N_0 \frac{\lambda_m \lambda_c}{\lambda_c - \lambda_m} \left(\frac{\exp\left[(\lambda_f - \lambda_m)t\right]}{\lambda_f - \lambda_m} - \frac{\exp\left[(\lambda_f - \lambda_c)t\right]}{\lambda_f - \lambda_c} + \frac{\lambda_c - \lambda_m}{(\lambda_f - \lambda_c)(\lambda_f - \lambda_m)} \right) \tag{3.16}$$

Then, after some simplification:

$$\begin{aligned} \phi_2(t) &= \frac{(\lambda_c - \lambda_m)(\lambda_f - \lambda_m)(\lambda_f - \lambda_c)}{\lambda_m \lambda_c} \\ &\quad \cdot \frac{\phi_1(t)}{(\lambda_f - \lambda_c)\exp[-\lambda_m t] - (\lambda_f - \lambda_m)\exp[-\lambda_c t] + (\lambda_c - \lambda_m)\exp[-\lambda_f t]} \end{aligned} \tag{3.17}$$

Parameter estimation

Values of λ_m, λ_f, λ_c and T were estimated by minimising the error function:

$$\begin{aligned} &\epsilon_1(\lambda_m, \lambda_c, \lambda_f, T) = \\ &\sum_i \left(\left[\widehat{\phi}_1^i(t_i) - \phi_1(t_i, \lambda_m, \lambda_c, \lambda_f, T) \right]^2 + \left[\widehat{\phi}_2^i(t_i) - \phi_2(t_i, \lambda_m, \lambda_c, \lambda_f, T) \right]^2 \right) \end{aligned} \tag{3.18}$$

where $\widehat{\phi}_1^i(t_i)$ and $\widehat{\phi}_2^i(t_i)$ refer to the data collected on patient i, aged t_i. Minimisation was carried out in R using the default Nealder-Meld simplex method (R DEVELOPMENT CORE TEAM, 2007).

Confidence intervals were computed using a case resampling bootstrap (RICE, 1994). For each bootstrap, a new data set $\widehat{D}$ was constructed by sampling, with replacement, 62 samples from the original data of 62 patients. The value of $\epsilon_1(\lambda_m, \lambda_c, \lambda_f, T)$ was then minimised using the data set $\widehat{D}$, and the computed values of λ_f, λ_m, λ_c and T were recorded. Repeated bootstraps ($\geq 10^4$ repeats) gave approximations to the underlying sampling distributions of the parameters. Confidence intervals for the parameter estimates were computed by calculating the 2.5 and 97.5 percentiles of the sampling distributions.

Estimated parameter values, with bootstrap confidence intervals, are displayed in Table 3.2.

3.4.5 Model 2: Definition

Model 2 was a simplified version of Model 1 that neglected to describe monoclonal conversion. In Model 2, wholly mutated crypts were produced from wild-type crypts at a rate of λ_m per crypt per year. The equations for this modified model were:

$$\frac{dN}{dt} = \lambda_f N - \lambda_m N \tag{3.19}$$
$$\frac{dM}{dt} = \lambda_m N + \lambda_f M \tag{3.20}$$
$$\frac{dP}{dt} = \lambda_m N \tag{3.21}$$
$$\frac{d\tau}{dt} = \lambda_f \tau \tag{3.22}$$

with the initial conditions:

$$\tau(0) = N(0) = N_0; M(0) = P(0) = 0 \tag{3.23}$$

As before, the total number of mutant crypts was computed as the total number of crypts minus the number of wild-type crypts:

$$M(t) = \tau(t) - N(t) \tag{3.24}$$

Integrating equations (3.19) and (3.22) gave:

$$N(t) = N_0 \exp\left[(\lambda_f - \lambda_m)t\right] \tag{3.25}$$

and

$$\tau(t) = N_0 \exp\left[\lambda_f t\right] \tag{3.26}$$

Substituting these results into (3.24) gave an expression for $M(t)$:

$$M(t) = N_0 \exp[\lambda_f t]\left(1 - \exp[-\lambda_m t]\right) \tag{3.27}$$

Finally, integrating (3.21) gave:

$$P(t) = N_0 \frac{\lambda_m}{\lambda_f - \lambda_m} \exp[\lambda_f t]\left(\exp[-\lambda_m t] - \exp[-\lambda_f t]\right) \tag{3.28}$$

The two statistics $\phi_1(t)$, the proportion of mutated crypts at time t, and $\phi_2(t)$ the mean patch size at time t, were used to fit the model to the data.

By definition:

$$\phi_1(t) = \frac{M(t)}{\tau(t)} = 1 - \exp[-\lambda_m t] \tag{3.29}$$

$$\phi_2(t) = \frac{M(t)}{P(t)} = \frac{\lambda_f - \lambda_m}{\lambda_m} \frac{1 - \exp[-\lambda_m t]}{\exp[-\lambda_m t] - \exp[-\lambda_f t]} \tag{3.30}$$

Estimate of λ_m and T

Taking logs of equation (3.29) gives:

$$\log[1 - \phi_1(t)] = -\lambda_m t \tag{3.31}$$

And since $t = t' - T$, this gives:

$$\log[1 - \phi_1(t)] = -\lambda_m t' + \lambda_m T \tag{3.32}$$

which is of the form $y = mx + c$. Fitting (3.32) to the data $\{\widehat{\phi_1^i}(t_i)\}$, using the default least-squares method in R (R DEVELOPMENT CORE TEAM, 2007), provided an estimate of T as $T = c/m$.

Estimate of λ_f

By definition, for $t' \leq T$ ($t \leq 0$), $\phi_2(t) = 1$. For $t' > T$ ($t > 0$), $\phi_2(t)$ is as specified in (3.29). The value of λ_f was estimated by minimising the square error:

$$\epsilon_1(\lambda_f) = \sum_i \left[\widehat{\phi_2^i}(t_i) - \phi_2(t_i, \lambda_f)\right]^2 \tag{3.33}$$

using the value of λ_m was computed in 3.4.5. As before, $\widehat{\phi_2^i}$ refers to the data for patient i. Confidence intervals were computed using the bootstrap approach described above.

3.4.6 Stochastic model of mutant clone expansion

To investigate how crypt fission affected the fixation of mutant cells in the colon it was necessary to describe the dynamics of stem cells in the niche. To do this, the model of Yatabe *et al.* was adapted (for full model definition see supplementary materials of

YATABE *et al.* (2001)). In their model, Yatabe and colleagues defined the probabilities p_0, p_1 and p_2 of a stem cell having 0, 1 or 2 stem cell offspring in a niche that contains a constant N stem cells. The N stem cells were assumed to divide synchronously, at rate of one division every 48 hours. Crypt sizes are normally distributed (TOTAFURNO *et al.*, 1987) so it was assumed that a crypt containing N stem cells represented the mean stem cell pool size. Division probabilities were constrained so that precisely N stem cells remained in the niche following fission. Yatabe *et al.* estimated the value of p_0, p_1 and p_2 by comparing their model to data on methylation at neutral loci (see Introduction, Section 1.4.1), and suggested that $p_1 = 0.95$, and $p_0 = p_2 = 0.025$. Yatabe and colleagues also described the differentiated cell population in their model; here non-stem cells were assumed to remain in the crypt for only a short time, and so were neglected.

In this Chapter, the model of Yatabe *et at* was modified to describe crypt fission. At the initiation of crypt fission, stem cells were divided randomly between the two daughter crypts. "Symmetrical fission", whereby the N stem cells were divided evenly between the two daughter crypts, and "asymmetrical fission" which resulted in non-even segregation of stem cells, were considered. Without loss of generality, the number of the original stem cells that were found in the first daughter crypt (n_1) after division was $n_1 = 1 + \text{binomial}(N-2, \rho)$. The number of parent cells in the second daughter crypt was then $n_2 = N - n_1$. For symmetrical fission, $\rho = 0.5$, and increasingly asymmetrical fission was represented by decreasing ρ. After fission, the full complement of stem cells in the daughter crypts was restored by selecting at random an appropriate number of stem cells to divide symmetrically to produce two SC offspring. In this manner, crypt fission could lead to a partially-mutated crypt becoming wholly-mutated. For example, a partially mutated crypt could undergo fission in such a way that one of the daughter crypts contained only COX-deficient cells. It was assumed that COX-deficient cells were randomly distributed between the two daughter crypts. Given that a parent crypt contained M COX-deficient stem cells before fission, the number of COX-deficient stem cells found in the first daughter crypt after fission (k) followed a hypergeometric distribution with parameters N, M and n_1, where n_1 is the number of parent stem cells that are segregated to the first daughter crypt during

fission.

Simulations began with a crypt that contained a single COX-deficient stem cell. The niche dynamics model for the crypt was then run, with a fission rate of λ_f and fission symmetry parameter ρ, until the COX-deficient clone was lost from all crypts, or until a crypt containing only COX-deficient stem cells was formed. The probability that a mutant cell would succeed in the niche after a given time, with a fission rate of λ_f, was approximated by repeatedly running the model and recording the proportion of runs where the mutant stem cell succeeded in the niche, or was lost from the niche.

3.5 Results

3.5.1 Crypt doubling time

The rate of crypt fission events, λ_f per crypt per year, was estimated as described above. λ_f had to be co-estimated with λ_m, the formation rate of COX-deficient cells, λ_c, the rate at which partially mutated crypts became wholly mutated, and T, the age at which the first COX-deficient cells formed.

Estimated values of the parameters for Model 1 and Model 2 are displayed in Tables 3.2 and 3.3 respectively. The estimated value of λ_f in Model 1 was 0.042 (CI: 0.031-0.058). This corresponded to a mean crypt cycle duration of 23.6 years (CI: 17.3-32.6 years). The fit of Model 1 to the data is displayed in Figure 3.3.

The estimated value of λ_f from Model 2 ($\lambda_f = 0.042$, CI: 0.024-0.058) compared favourably to Model 1 ($\lambda_f = 0.042$, CI: 0.031-0.058), indicating that the value of λ_c (in Model 1) was indeed larger than the other rate parameters and did not significantly influence the estimation of λ_f. Correspondingly, the 95% confidence intervals for λ_c were very wide, as illustrated in Figure 3.4, again illustrating that the estimate of the rate of crypt fission did not depend significantly on the rate of monoclonal conversion.

The estimated value of T was around 35 years in both models, consistent with the observation that few COX-deficient crypts are observed in patients aged less than 40 years.

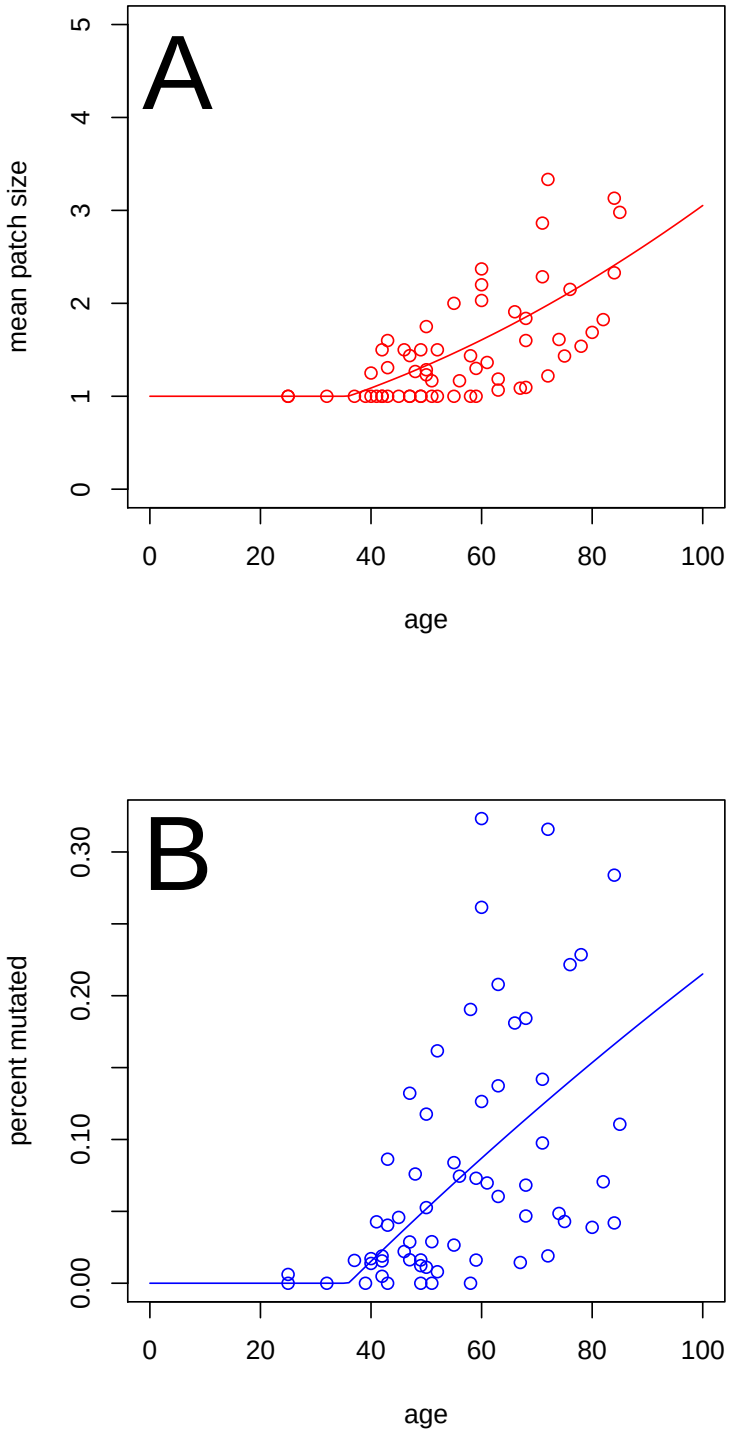


Figure 3.3: Fit of Model 1 to mean patch size and proportion of mutated crypts data. **A:** Mean patch size vs age. Mean patch size of zero implies no mutated crypts were observed in the colon biopsy. **B:** % of colon mutated vs age. The model was computed with the estimated parameter values as displayed in Table 3.2.

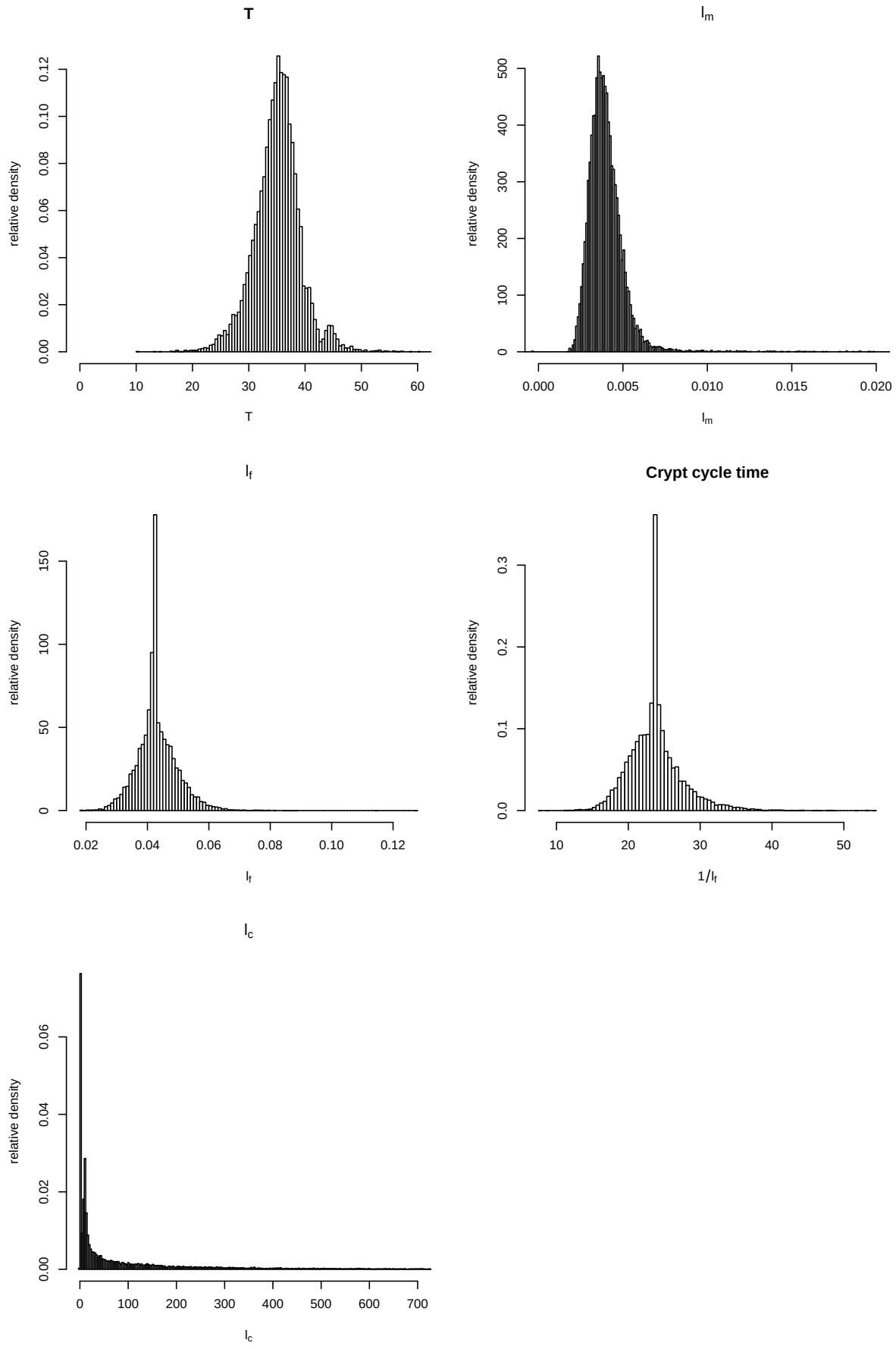


Figure 3.4: Histograms showing bootstrapped parameter estimates for Model 1. From left to right, top to bottom: bootstrapped values of T, λ_m, λ_f, crypt cycle time, and λ_c.

Table 3.2: Estimated parameter values for Model 1

Parameter	Estimate	Median	95% limits of estimated parameters
T	35.8	35.2	(26.3-44.2)
λ_m	0.004	0.004	(0.003-0.006)
λ_f	0.042	0.042	(0.031-0.058)
$1/\lambda_f$	23.6	23.6	(17.3-32.6)
λ_c	6.9	23.3	(0.054-1708.6)

Table 3.3: Estimated parameter values for Model 2

Parameter	Estimate	Median	95% limits of estimated parameters
T	35.5	34.8	(18.1-41.8)
λ_m	0.003	0.003	(0.001-0.007)
λ_f	0.042	0.041	(0.024-0.058)
$1/\lambda_f$	23.8	24.6	(17.2-42.4)

3.5.2 Fixation of mutations

Using the stochastic model described above, the rate at which fission fixed mutations in the epithelium was examined. Mutations were assumed to have no selective benefit or detriment in the stem cell niche. The fixation of mutations due to fission depended upon two parameters: ρ, which controlled the probability of even distribution of stem cells between the two daughter crypts, and λ_f, the rate of crypt fission.

A tendency for "symmetrical crypt fission" ($\rho \simeq 0.5$), such that each daughter crypt was likely to receive half the total number of stem cells in the parent crypt, increased the likelihood that a mutation would eventually become fixed in a crypt (see Figure 3.5). When $\rho = 0.5$, both daughter crypts were most likely to receive $N/2$ of the parent stem cells. To reach the total complement of N stem cells in each daughter crypt, each of the parent cells underwent a symmetric division which caused the number of

mutant cells to double. This increased the numbers of mutant cells, and meant that eventual fixation of the mutant clone was more likely. Immediately after the formation of a mutant clone, the likelihood of fixation of the clone was highest for values of $\rho \simeq 0$ (see Figure 3.5.2). This was because for small ρ, shortly after clone formation, daughter crypts could be formed which contained only a single parent stem cell. The full complement of the crypt was restored by repeated divisions of this founder cell. If the founder cell happened to be a mutant cell, this would result in the crypt formed entirely of mutant cells. For larger values of ρ, this *crypt purification* through fission was very unlikely to occur.

Mutant stem cells were increasingly likely to become fixed in a crypt (displace all wild-type cells) as the rate of fission increased (see Figure 3.5). The rate at which a clone becomes fixed in a crypt increased as the fission rate increased (see Figure 3.5.2). This is because fission causes an increase in stem cell number, and so an increased fission rate tends to increase the number of mutant cells which makes fixation of a mutant clone more likely.

3.6 Discussion

In this chapter, the duration of the crypt cycle was estimated at 23.6 years (CI: 17.3-32.6). The time taken for the number of crypts in the colon to double, termed the crypt doubling time (C) by Bjerknes and Cheng (CHENG *et al.*, 1986a; TOTAFURNO *et al.*, 1987) was $C = \ln(2)/\lambda_f$. Using the estimate of the crypt fission rate estimated in the first model presented in this chapter, the estimated crypt doubling time is 16.4 years (CI: 12.0 - 22.6), slightly longer than the previous estimate of 9-18 years. This is the first estimate of the crypt fission rate in the normal human colon inferred directly from human data.

Supposing that an individual lives to around 75 years of age, each original crypt will undergo approximately three fission events in the lifetime of the individual, implying that the number of crypts will increase eight-fold in an average human lifetime. How is this large increase in the number of crypts manifested in the appearance of the colon? BHATNAGAR *et al.* (2004) and HOUNNOU *et al.* (2002) looked at changes in the length and width of the colon with age, and observed that there was no significant fluctuation

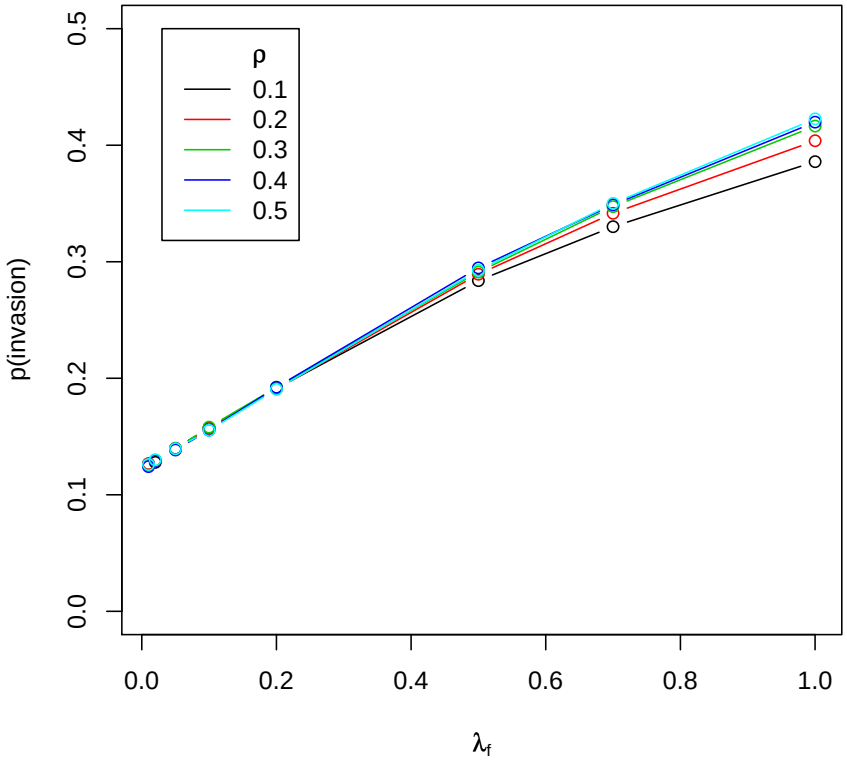


Figure 3.5: Probability of eventual mutant clone fixation as a function of the crypt fission rate λ_f and degree of symmetry in crypt fission ρ. A mutant lineage was more likely to become fixed in the epithelium as the crypt fission rate increased. As the tendency for a symmetric crypt fission event increased (increasing ρ), the likelihood of eventual mutant fixation also increased. Increasing the crypt fission rate had a more marked effect than increasing the likelihood of symmetric fission events. Eventual fixation was defined as fixation 20 years post clone formation. Simulations results are shown for $N = 8$ stem cells per crypt, 182 rounds of cell divisions per year.

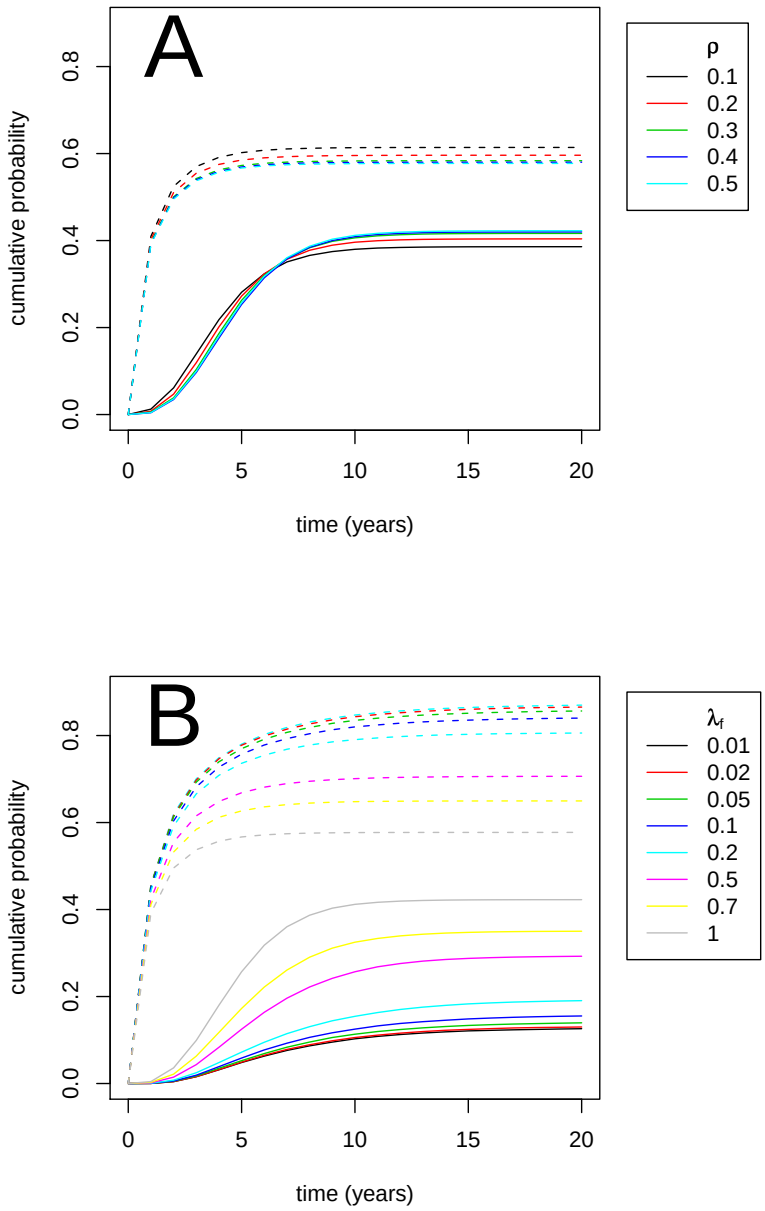


Figure 3.6: Rate of mutant clone invasion and loss for different values of the crypt fission rate and degree of symmetry in crypt fission. Dashed lines represent the probability of clone loss, solid lines the rate of clone invasion. **A:** Rate of mutant clone invasion and loss for various values of the crypt fission symmetry parameter ρ. Increasing ρ increased the likelihood of eventual fixation of a clone. Shortly after clone induction, fixation of the mutant clone was most likely for small ρ. **B:** Rate of mutant clone invasion and loss for various rates of crypt fission λ_f. A clone was more likely to reach fixation, and tended to reach fixation faster, as the rate of crypt fission increased. Simulations results are shown for $N = 8$ stem cells per crypt, 182 rounds of cell divisions per year, $\rho = 0.5$, and $\lambda_f = 1$ per year unless otherwise stated.

in size with age. Thus, if the number of crypts in the colon does indeed increase eight-fold over an average human lifetime, then this change is not manifested by the growth of a bigger colon. How then, could the colon accommodate a larger number of crypts? Casual observations suggest that colon becomes increasingly folded as it ages. The folds increase the surface area and so more crypts can be accommodated. Further, casual observations suggest that crypts are more densely packed in older people; the average crypt diameter may be reduced and so more crypts fit in a smaller area of the colonic epithelium. If it is supposed that, in early life, each crypt has an average diameter of r micrometres, then the area occupied by n crypts is proportional to nr^2. Thus, slight decreases in the mean crypt diameter permit many more crypts to fit in the same area of epithelium. For example, if the mean diameter of the n crypts were to halve, the area of the epithelium occupied by the crypts would be reduced to a quarter of the original area. The suggestions that the colon shows increased folding and the mean crypt size reduces as a person ages are speculative but merit investigation (see Section 3.8.1).

Alternatively, the absolute number of crypts in the colon may remain roughly constant with age. The rate of "crypt birth", through fission of parent crypts, may be balanced by an equal rate of *crypt death*. Crypt death, however, has not been observed: no crypts that appear to be in process of "crypt apoptosis" have been identified. In the murine colon, BJERKNES (1986) looked for evidence of crypt death. Sections of colon from 3 mice were repeatedly observed over a period of three weeks, by exorcising the colon, and no crypts were lost during the period of observation. However, the reliability of these findings are doubtful, as it is practically very difficult to confirm the identity of a crypt, as reference points such as tissue folds or blood vessels do not provide consistent markers. Further crypt replacement would have gone unnoticed. Given the large number of crypts in the colon, around 10^7, it seems inconceivable that there is not occasional crypt death. Further, it is reasonable to speculate that certain regions of the colon are more susceptible to crypt death, due to environmental differences between regions of the colon. Thus, crypt expansion in one area of the colon, which causes the expansion of a COX-deficient crypt to form a patch, may be balanced by crypt death elsewhere in the colon.

The idea that crypts have a natural life cycle, analogous to the cell cycle, was first suggested by Bjerknes and Cheng in 1986 (CHENG *et al.*, 1986a), after they noted that crypts in fission tended to be larger than crypts that were not in fission. Fission was proposed as a mechanism that occurred in response to periodic fluctuations in the stem cell pool (TOTAFURNO *et al.*, 1987). This idea has become prevalent in the literature (LOEFFLER *et al.*, 1997; EDWARDS and CHAPMAN, 2007). However, the possibility remains that crypts may have no characteristic life time. Crypt fission may simply be a "healing" response to epithelial damage, analogous to skin growth after injury, whereby a crypt divides to replace a neighbour that is no longer viable due to some environmental or internal stress. Stem cell numbers may still fluctuate in crypts, and influence the timing of a fission event. The oscillations in stem cell number may not be periodic though, but rather be a response to another signal which causes the initiation of fission or cell growth in the colon. To corroborate this idea, many authors have noted that the frequency of crypts in fission is greatly increased in regions of inflammation (WANG, 2006; CHEN *et al.*, 2005; CHENG *et al.*, 2000). Inflammation is associated with increased levels of oxidative stress (GRISHAM, 1994; PAVLICK *et al.*, 2002; RACHMILEWITZ *et al.*, 1995; SINGER *et al.*, 1996) which would presumably increase the rate of crypt death, due to increased rates of cell death inside the crypts (KANNAN and JAIN, 2000). The presence of COX-deficient patches can be explained if a COX-deficient crypt formed adjacent to a region of local inflammation, so that crypt fission was induced in the COX-deficient crypt to replenish the depleted crypt pool. This model for the initiation of crypt fission could be tested by a suitable stochastic model of inflammation and crypt fission (see Section 3.8.5).

In this chapter, cytochrome-c-oxidase deficiency in cells has been used as a marker of clonality. COX-deficiency is undoubtedly an excellent lineage marker, as it occurs relatively infrequently and is readily detectable by routine methods. However, using COX-deficient cells as a model to study normal homoeostatic processes assumes that COX-deficient cells behave as wild-type cells do. There are three characteristic length scales at which selection for COX-deficiency could occur in the colon: in the mitochondrial pool of a single cell, between cells in a crypt, and between crypts in the colon.

At the sub-cellular level, mathematical modelling of mtDNA dynamics suggests that the clonal expansion of a neutral mutation in the mtDNA pool, through genetic drift, is probably a slow process (ELSON *et al.*, 2001). Simulations suggest that it can take around 50 years for 1% of cells to become COX-deficient given that new mtDNA strands are generated at a rate of 10^{-5} per mtDNA replication. Higher mutation rates lead to greater numbers of cells becoming COX-deficient earlier in life. The data analysed in this chapter shows that a COX-deficient crypt is observed in most biopsy sections from patients over the age of 40. This relatively high frequency of COX-deficient crypts in the colons of young adults suggests either that COX mutant mtDNA rapidly reaches homoplasmy in the mtDNA pool, and so has a selective advantage in the mitochondrial pool, or that the rate of COX mutations is relatively high. A precise estimate of the mtDNA error rate is required to determine whether a high mutation rate or selection for the mutated mtDNA generates the large number of COX-deficient crypts observed in the colon. Deep re-sequencing of the entire mtDNA genome in patients of varying ages could be used to accurately estimate somatic mtDNA mutation rate.

Selection may also occur at the cell level. By analysing the accumulation of methylation in the promoter region of a gene that is not expressed in the colon, YATABE *et al.* (2001) suggested that the niche-succession time, the time taken for a single cell clone to become dominant in a crypt through random drift, was around 8.2 years. Thus, given this slow niche succession rate, if COX-deficient cells had no selective benefit compared to their neighbours, many partially mutated crypts would have been observed. In order to investigate the expansion of COX-deficient cells in an otherwise wild-type crypt, combining mtDNA and methylation analysis may be informative. The diversity of methylation patterns within the cells in a COX-deficient clone could be used to infer the expansion rate of the mutated clone.

Given that selection for COX-deficiency may well occur at the mtDNA and cell levels, do COX-deficient crypts also have a growth advantage compared to their wild-type counterparts? Determining this is crucial if the estimate of the crypt fission rate derived in this Chapter is to be used as the estimate of wild-type colonic crypts. Selection at the crypt level is difficult to infer, since no comparison can be made with the expansion

rate of non-COX-deficient crypts. If COX-deficient crypts did have a growth advantage compared to wild-type crypts, patches of COX-deficient crypts would be larger than patches of related wild-type crypts. Correspondingly, the duration of the crypt cycle in wild-type crypts would be longer than estimated in this chapter.

Crypt fission increases the likelihood that a mutation will become *fixed* (see Section 3.5.2), and further fission events facilitate the spread of mutations through the epithelium (GREAVES *et al.*, 2006). The low fission rate estimated in this Chapter, of a fission event in each crypt once every 23.6 years (CI: 17.3-32.6 years), is perhaps an adaptive anti-tumourigenic mechanism which prevents the spread of mutations through the GI tract. The primary expansion of colorectal adenomas is through crypt fission (PRESTON *et al.*, 2003). The genetic event(s) which cause the increase in the crypt fission rate must be a crucial early step in the development of a colorectal adenoma.

The number of crypts in fission is significantly increased in adenomas compared to the wild-type epithelium (WONG *et al.*, 2002). A sixteen-fold increase in the number of crypts in fission was observed by PRESTON *et al.* (2003) in early sporadic adenomas (3% of crypts in fission in normal mucosa, 48% of crypts in the adenoma), which confirmed the earlier observations of WONG *et al.* (2002). If the sixteen-fold increase in the number of crypts in fission in an adenoma is indicative of a sixteen-fold increase in the fission rate, the duration of the crypt cycle would be reduced to 1.5 years (CI 1.1 - 2.0 years) in colorectal adenomas. The rate of fission may be much higher than this if the time taken for each fission event is also decreased in adenomas.

The frequency of crypts in fission is increased in patients with FAP (WASAN *et al.*, 1998; BJERKNES *et al.*, 1997). These patients carry a heterozygous germline mutation in the APC gene (see Introduction, Section 1.6.1), which suggests that APC maybe involved in the regulation of crypt fission. This involvement of APC could be direct, for example by stimulating or repressing a particular genetic pathway that regulates crypt fission. More likely though, is that the action of APC is indirect. Phenotypic characteristics of a crypt, that are under the control of APC, may in turn cause the initiation of fission. Mathematical modelling by BOMAN *et al.* (2001) suggests that the number of stem cells per crypt is increased in FAP patients, and similar conclusions

have been drawn from the analysis of methylation patterns by KIM *et al.* (2004c). Crypts in fission tend to be larger than average (BJERKNES, 1986; WONG *et al.*, 2002), so are likely to contain a greater number of stem cells. Thus the frequency of crypt fission events in FAP colons may be a result of an increased number of stem cells per crypt. This hypothesis suggests that the single remaining wild-type copy of APC is insufficient to regulate crypt cells in a wild-type fashion. However in contrast to this theory, BJERKNES *et al.* (1997) suggested that the increased frequency of crypts in fission in FAP colons was due only to an increased number of crypts in fission in aberrant crypt foci (ACF). Given that, in FAP colons, mono-cryptal adenomas tend to have bi-allelic APC mutations (TAKAYAMA *et al.*, 2001), the increase in the crypt fission rate in FAP colons would not then be due to haploinsufficiency. The increase in the number of crypts in fission in FAP colons and adenomas could also be indicative of an increase in the time taken for a crypt to undergo fission, or even of a systemic failure to complete crypt fission. Crypts from FAP colons that are undergoing fission often show multiple "buds", each of which may be a precursor to a new crypt, so it appears that in colons new fission events can occur before previous crypt divisions have finished.

Fission appears to take a reasonably long time to complete in the normal colon: by using methylation at non-expressed promoters as a lineage marker, KIM and SHIBATA (2004) found that cells from the two arms of a bifurcating crypt were no more closely related than cells from two non-adjacent crypts. This suggests that the time taken for a crypt to divide is long enough to allow the methylation patterns of the cells in each arm of the dividing crypt to diverge. This work could be extended to estimate the duration of a fission event. A mathematical model could be developed which predicted the expected degree of dissimilarity between the methylation patterns from cells from the two arms of the bifurcating crypt, given that the fission event takes τ years to complete (see Section 3.8.2). Comparing the model to data on the methylation patterns on bifurcating crypts would allow τ to be estimated. This analysis could then be applied to determine whether the observed increase in the number of crypts in fission in an adenoma is due to an increase in the fission rate or a failure to complete fission.

3.7 Conclusion

Crypt fission is a rare event in the normal human colon. In this Chapter, it was estimated that each crypt divides once every 23.6 years (CI: 17.3-32.6 years). Fission increases the likelihood that a mutant stem cell will become fixed in a crypt, and facilitates the expansion of mutations through the epithelium. Fission is the primary method of expansion of colorectal neoplasms. Therefore understanding the mechanism responsible for increasing the fission rate is necessary for understanding the mechanism of tumorigenesis.

3.8 Future work

3.8.1 Changes in colon with age

The number of crypts in the colon may fluctuate with age. Mechanisms that may facilitate the colon to house a greater number of crypts are: increased folding of the colonic epithelium, or a decrease in the mean crypt size. These factors should be examined. Sections of colon biopsies of uniform area should be examined from a large number of patients of different ages. The number of crypts in each unit area should be counted. These counts should then be examined to see if the mean number of crypts per unit area depends upon the age of the patient at biopsy. Sections should also be examined for folds to see if the number of folds increases with age.

3.8.2 The duration of a fission event

The duration of a crypt fission event is another parameter of interest, which may be altered in neoplastic development. The duration of a fission event could be estimated by examining the diversity of methylation patterns in crypts that were undergoing fission. Methylation of CpG islands in the promoters of non-expressed genes has been shown to be a high fidelity lineage marker, which can be used to distinguish closely related cell lineages (YATABE *et al.*, 2001; KIM *et al.*, 2004c; KIM and SHIBATA, 2004; KIM *et al.*, 2005d,c,b, 2006; CHU *et al.*, 2007). A stochastic model of the change in diversity of methylation patterns during a fission event should be developed (see Figure 3.7). The model should be compared to the methylation patterns observed in crypts in fission.

This comparison would allow the duration of a crypt fission event to be inferred.

An outline of a suggested model is as follows: suppose the crypt in fission was removed from the colon of a patient of age A. The parameter of interest in the model is τ, the duration of a fission event. At the time when the crypt was removed, the crypt was part way through the fission event. Denote this time by δ, where $\delta < \tau$. Thus the fission event began at age $A - \delta$. δ is a nuisance parameter in the model, which follows some distribution on $[0, \tau)$.

The diversity of methylation patterns in a crypt is described by a statistic S. Using a model of stem cell dynamics and crypt fission (such as described above in Section 3.4.6), coupled with a model of methylation (such as that of YATABE *et al.* (2001)), the value of S in a crypt which began fission at time $A - \delta$ could be simulated. A rejection method could then be used to estimate τ. For each run of the model, a value of $\widehat{\tau}$ should be sampled from a reasonable prior distribution, then a corresponding value of δ should be sampled from the appropriate distribution on $[0, \tau)$. The value of the statistic S at time A could then be computed using these parameter values. $\widehat{\tau}$ would be accepted into the posterior distribution if the computed value of S was sufficiently similar to the value of S observed in a crypt in fission.

3.8.3 Inflammation induced crypt fission

Crypt fission occurs in response to inflammation in the colon (WANG, 2006; CHEN *et al.*, 2005; CHENG *et al.*, 2000). Therefore, crypts may not have a characteristic life-cycle duration, but instead divide in response to local inflammation. To examine the plausibility of this hypothesis, predictions from a suitable stochastic model of inflammation-driven crypt fission could be compared to the data on COX-deficient patches of crypts presented in this chapter. Inflammation driven crypt fission may better explain the large variation in patch size and frequency seen in different patients.

3.8.4 Modelling mtDNA dynamics

For a cell to have detectable COX-deficiency, a significant proportion of the mtDNA pool inside the cell must carry the mutation in the COX gene. COX mutations occur frequently in colonic crypts (see above), thus it appears that the expansion of a COX^-

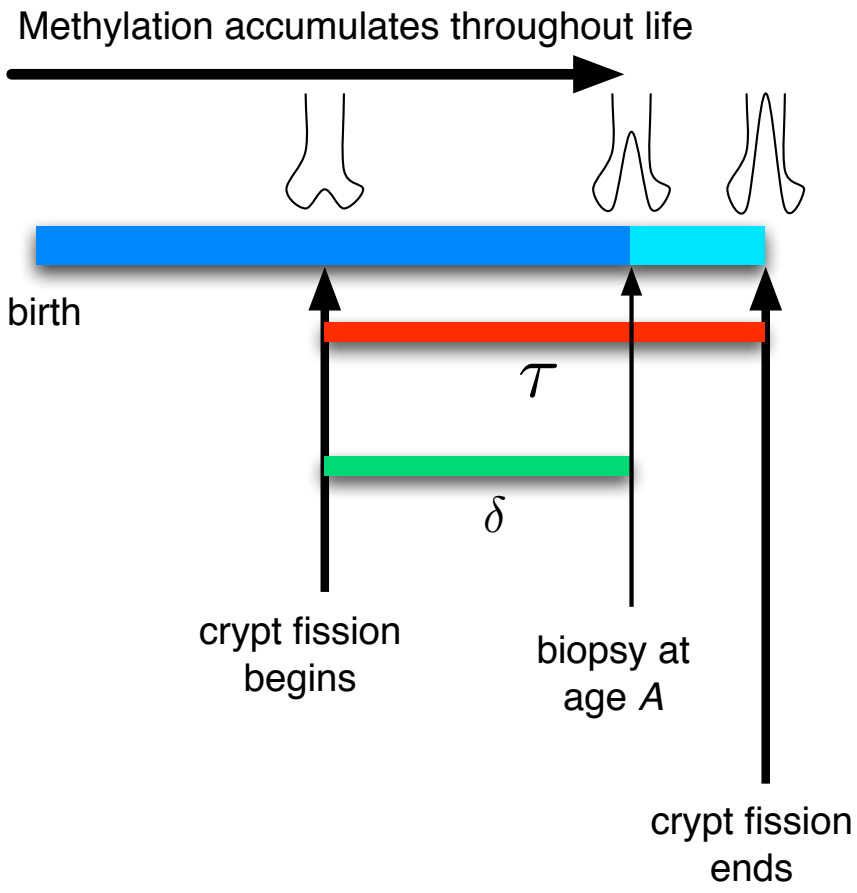


Figure 3.7: Model to estimate the duration of a crypt fission event. τ is the duration a fission event, indicated by the red line, and is the parameter of interest. δ, the time since fission began, is a nuisance parameter, and is denoted by indicated by the green line. Methylation accumulates in the crypt from birth. The diversity of methylation patterns in a crypt that is part way through fission could be compared to the model to estimate τ.

mtDNA strand inside a cell occurs rapidly. Mathematical modelling of the clonal expansion of the COX^- mtDNA might allow the selective advantage conferred by the mutation in the mtDNA pool to be inferred. Also, the rate at which COX mtDNA mutations occur could be estimated. Published models of mtDNA dynamics, such as those of ELSON *et al.* (2001) and COLLER *et al.* (2001) could be adapted for this purpose.

3.8.5 Stochastic model to estimate crypt fission rate

In this chapter, a deterministic model of the mutation and clonal expansion of COX-deficient cells was used to estimate the frequency of crypt fission events. This data may be better described by a stochastic model where the duration of crypt fission follows some underlying probability distribution. This could be incorporated into a model. The timing of the first COX mutation is likely to be a stochastic event, and this could be described in the model.

Chapter 4

Laboratory methods

4.1 Aims

- Establish tissue preparation protocol.
- Establish macro-dissection protocol.
- Establish laser capture micro-dissection protocol for single crypt dissection.
- Optimise sequencing methods for DNA extracted from a single crypt.

4.2 Introduction

The laboratory experiments presented in this thesis were designed to examine genetic heterogeneity in a colorectal adenoma or cancer. Most previous work on colorectal cancers has involved needle macro-dissection of a broad section of tumour tissue (see for example GONZÁLEZ-GARCÍA *et al.* (2002); JONES *et al.* (2007)). Although studies like these could distinguish macro-scale heterogeneity, such as genetic differences between the left side and the right side of the tumour, determining micro-scale heterogeneity, such as difference between neighbouring crypts, was not possible. Micro-scale descriptions of the genetic composition of an adenoma, termed a *clonal map*, are necessary to examine the development of tumour sub-clones. Genotyping micro-dissected tissue from formalin fixed paraffin embedded samples is technically difficult. Micro-dissection of single crypts yields only small quantities of DNA, which is also often of poor quality. Highly optimised genotyping techniques are required to

analyse adenomas on a crypt-by-crypt basis.

In this chapter, methods for the dissection and genotyping of individual crypts from paraffin embedded tissue are presented. Refined laser-capture micro-dissection techniques are described. Optimised genotyping methods for detection of APC and K-ras mutations in individual crypts are presented.

The methods described here were developed with Simon Leedham and Christina Thirwell at the Cancer Research UK London Research Institute.

4.3 Histology

4.3.1 Slide preparation

Formalin fixed, paraffin embedded tissue was used in this study. Ethical permission to use this tissue was granted by the Oxfordshire Research Ethics Committee B.

De-waxing

Slides were de-waxed by incubation in sulphur free xylene (VWR, UK). After de-waxing, tissue was rehydrated by sequential incubations in decreasing concentrations of ethanol (100%, 95%, 80%, 70% and 50%) and finally in distilled water.

4.3.2 H&E

Tissue histology was assessed with the aid of *Hematoxylin and Eosin* staining (H&E stain). *Hematoxylin*, a basic dye, colours acidic structures, such as the nucleus and ribosomes, with a blue-purple hue. *Eosin*, an acidic dye, colours basic structures, such as the cell cytoplasm, bright pink or red.

For H&E staining, slides were first de-waxed according to the protocol above. Harris Hematoxylin and Eosin stains (VWR, UK) were applied to the de-waxed slides, and then stained tissue was rehydrated as described above. Prepared slides were then mounted in DPX (VWR, UK). De-waxing and H&E preparation was automated using a Tissue-Tek Prisma automated slide stainer machine (Sakura, USA).

4.3.3 Dysplasia

Dysplastic tissue was identified by examination of the H&E slides. Dysplastic crypts could be identified by malformed crypt structure, multiple layers of cells in the crypt, and increased nuclear (Hematoxylin) staining. Dysplasia identification was confirmed by at least one expert pathologist (Dr Maesha Deheragoda or Prof. Nicholas Wright).

4.4 Dissection and DNA Extraction

Both macro and micro dissection were performed.

4.4.1 Macro-dissection

Macro-dissection digestion buffer

Digestion buffer for macro-dissection was made from 4.45ml distilled H_20, 500μl 10x magnesium-free PCR buffer, at pH8 with 50μl proteinase K (VWR, UK) at 200μg/ml. Between, 30-100μl of digestion buffer was used per sample, depending on the mass of tissue being digested.

Macro-dissection process

Paraffin embedded tissue was cut onto non-coated glass slides (Thermo Shandon Colorfrost), at thickness of 6μm using a Jung Biocut 2035 microtome. Tissue was typically cut into six serial sections, although fewer slides were used for larger adenomas. Slides were de-waxed as described above (Section 4.3.1). De-waxed tissue was then lightly stained with tolonium chloride (0.2% ppm, VWR, UK) to aid identification of regions of interest. Stained tissue was then air dryed in an incubator at 37°C for approximately one hour.

Macro-dissection was then performed using a sterile needle under a dissecting microscope (Nikon Eclipse E400, Nikon Europe, The Netherlands). The needle was dipped in the digestion buffer, as the buffer aided detachment of the tissue from the slide. Tissue was transferred to a sterile Eppendorf tube containing an appropriate volume of digestion buffer for the mass of tissue dissected, typically a 0.5cm^2 area of

tissue would be digested in 50μml of digestion buffer. Dissected tissue was transferred to a 55°C incubator (Techne Hybridsier HB-1D, Techne, UK) for a minimum of 24 hours and agitated frequently (RotaMixer, Hook and Tucker Instruments, UK). After incubation, the proteinase K was deactivated by heating to 95°C for 10 minutes and then centrifuged at 12000 rpm for 15 minutes (Eppendorf centrifuge 5154 D). The DNA rich supernatant was removed, and the protein rich precipitate was discarded. DNA was stored by freezing at-18°C or lower, or if required for immediate use, was held in a standard refrigerator at 2-8°C.

4.4.2 Micro-dissection

Laser capture micro-dissection

Micro-dissection was perform with the aid of a laser capture dissecting microscope (PALM, Carl Zeiss). Laser capture micro-dissection permits extremely precise dissection, allowing homogeneous cell populations can be collected (ESPINA *et al.*, 2006). Here, laser capture micro-dissection was used to extract specific regions of tissue, usually single crypts. This dissection technique is performed robotically, so the risk of contamination is reduced.

Laser capture micro-dissection required the use of PALM Membrane Slides (PALM micro-laser technologies AG, Germany), which had a thin plastic coating on one side of the slide. Cut tissue is transferred onto this plastic coating. The plastic film provides support for the tissue and prevents fragmentation during laser cutting and transfer of tissue from the slide to a capture tube. PALM slides were 254 nm wavelength UV light pre-treated for 30 minutes to improve section adhesion. Paraffin embedded tissue was cut at a thickness of 6μm onto PALM slides using a Jung Biocut 2035 microtome (Leica Instruments, Germany). The tissue was cut as ordered serial sections according to the following protocol: sections 1-3 were the PALM slides, section 4 was H&E stained to provide a histological reference, and slides 5-7 were PALM slides. Additionally, sections 0 and 8 were occasionally made into additional H&E slides to aid histology identification.

PALM slides were dried overnight at 37°C (Windsor Incubator, Windsor, UK). After

drying, slides were de-waxed according to the de-waxing protocol described above (Section 4.4.1). After de-waxing, slides were lightly stained with methyl green (0.1% ppm, VWR, UK), and dried at 37°C for at least 1 hour. A PALM laser micro-dissection microscope and associated PALM Robo software (PALM, Carl Zeiss, UK) was used to perform micro-dissection. Using the microscope, the H&E was examined to identify a crypt, or part crypt, of interest, then each of the six PALM slides were examined to confirm that the same crypt could be identified in each slide. The crypt was cut from each section, and cut tissue was catapulted into a sterile collection tube.

Digestion of micro-dissected tissue

A single crypt was digested in 10μl of proteinase K digestion buffer (PicoPure DNA Extraction Kit, Arcturus Bioscience) incubated at 65°C (Techne Hybridsier HB-1D, Techne, UK) for a minimum of twelve hours. Prior to digestion, proteinase K digestion buffer was kept on ice to optimise digestion yield. An aliquot of pure digestion buffer was also digested, and was subsequent used as to confirm that the buffer was not contaminated. After digestion, proteinase K was deactivated by heating to 95°C for 10 minutes and then spun at 5000 rpm for 3 minutes (Eppendorf centrifuge 5154 D). A small volume of distilled water, typically 5μl, was added to the sample to replace the buffer lost due to evaporation during the digestion and deactivation process.

4.5 APC mutation screening

Tumorigenic mutations in APC tend to cluster in the region between codons 1250-1450, termed the mutator cluster region (SEGDITSAS and TOMLINSON, 2006). The MCR has been divided into 12 overlapping regions, each of around 100-200 base pairs, to facilitate PCR amplification (see Appendix Section 8.2). Sequencing of APC was performed on the small quantity of DNA extracted from a single crypt. Due to the small quantity of DNA available, direct sequencing of the entire MCR of each crypt was impractical. Hence, prior to direct sequencing, single strand conformation polymorphism (SSCP) was performed on macro-dissected tissue to identify which of the 12 MCR sub-divisions harboured a mutation. Then only these mutated regions needed to be directly sequenced in each crypt.

4.5.1 Single strand conformation polymorphism

SSCP was used to screen for mutated regions of the APC gene (HAYASHI, 1991). To perform SSCP, the genomic region of interest was amplified in both tumour and control samples, using standard PCR with fluorescently labelled primers. PCR products were then denatured, and the resulting single strands of DNA naturally folded to form a tertiary structure. Mutations alter the conformation of this structure. Differences between samples could then be resolved by gel electrophoresis, or an analogous capillary system, since a sample's conformation determined its motility through the gel.

Method

Macro-dissection and DNA extraction was performed on whole polyp sections according to the protocol above (section 4.4.1). Twelve individual PCRs, one for each of the MCR regions, were then performed on the extracted DNA. Details of the primers for each MCR region are given in Table 4.1. Reagent quantities and reaction conditions for each MCR subdivision are shown in Table 4.2. Reagent concentrations and manufacturer's details are shown in Table 4.3. To confirm that PCR master mix was not contaminated, blank (master mix only) reactions were always performed. Reactions were performed in 96 well plate that had been sealed with adhesive PCR film (Thermo Scientific, UK). The reaction was carried on a Peltier Thermal Cycler (PTC-225, MJ Research, USA) according to the protocol in Table 4.13.

PCR efficiency was examined by gel electrophoresis (see Section 4.7). Concentrated reactions products, identified by the presence of a strong band on the electrophoresis gel, were diluted with double distilled H_20 to an appropriate concentration between 1:2 - 1:100. 4μl of diluted PCR products were aliquotted to a new 96 well plate. Amplified DNA was denatured with Hi-Di (Applied Biosystems, USA) according to manufacturer's instructions. An ABI 3100 16 capillary DNA analyser (Applied Biosystems, USA) was used to resolve the samples. Results were analysed in the Genotyper software (Applied Biosystems, Perkin Elmer), using the "view by scan" option. Mutated regions were identified by visually comparing tumour to wild-type DNA (see Figure 4.5.1).

Table 4.1: APC mutation cluster region SSCP primer details

Primer name	Forward primer	Reverse primer
MCR 1	CAAGCAGTGAGAATACGTCCA	TTTCTTGGTTAATAGAAGAAACTTTGC
MCR 2	GCCACTTGCAAAGTTTCTTCT	TGCTTCCTGTGTCGTCTGA
MCR 3	CAGACGACACAGGAAGCAGA	TGGAACTTCGCTCACAGGAT
MCR 4	GAAGATCCTGTGAGCGAAGTTCC	CAAGTCCTCTGGGGTGAGTA
MCR 5	AGAATCAGCCAGGCACAAAG	GCAATCGAACGACTCTCAAA
MCR 6	CACTATGTTCAGGAGACCCCA	TGGAAGATCACTGGGGCTTA
MCR 7	GTGAACCATGCAGTGGAATG	ACTTCTCGCTTGGTTTGAGC
MCR 8	GCTTAGGTCCACTCTCTCTCTT	GGCATTATAAGCCCCAGT
MCR 9	CCTAAAAATAAAGCACCTACTGCTG	CACTCAGGCTGGATGAACAA
MCR 10	ACATTTTGCCACGGAAAGTA	GGCTGCTCTGATTCTGTTTC
MCR 11	CAGGAAAATGACAATGGGAAT	TGGCATGGCAGAAATAATACA
MCR 12	GGACCTATTAGATGATTCAGATGATG	ACTTGGTTTCCTTGCCACAG

Table 4.2: APC mutation cluster region SSCP PCR reagent quantities and annealing temperatures. All volumes in μl and temperatures in °C. Volumes listed are for a single 25μl reaction.

	MCR											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Forward primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Reverse primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
dNTP	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Buffer	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Mg	1	2	1	1	1	1	1	1	3	2	2	2
H_20	18.25	12.25	18.25	18.25	18.25	18.25	18.25	18.25	16.25	17.25	17.25	17.25
Q		5										
Taq	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
DNA	2	2	2	2	2	2	2	2	2	2	2	2
Temperature	55	55	60	60	55	60	60	60	50	55	55	55

Table 4.3: PCR reagent concentrations and manufacturer's details.

Reagent	Manufacturer	Working Concentration
Primers	Sigma-Aldrich, UK	20mM
Taq	Qiagen, UK	5 units / μl
dNTP	Promega,UK	10 mM each dNTP
Mg^{2+}	Qiagen, UK	25mM
Buffer	Promega, UK	as supplied
Q	Qiagen	5x solution (as supplied)

Table 4.4: Thermal cycling protocol for PCR reactions.

Step	**Description**
1	94°C for 4 minutes
2	94°C for 1 minute
3	annealing temperature °C for 1 minute
4	72°C for 1 minute
5	Repeat steps 2-5 34 more times
6	72°C for 10 minutes
7	8°C forever (holding temperature)

4.5.2 Direct sequencing from laser-captured tissue

APC sequencing was performed on the very small volumes of DNA extracted from a single micro-dissected crypt. Nested PCR reactions, which used DNA from the first round PCR reaction as a template for the second, were used to produce sufficient DNA yield for sequencing. First round primers were designed to amplify a region containing both the MCR and the target sequence for the second round primers (see schematic in Figure 4.5.2). Primer details are shown in Tables 4.5 and 4.6, and the location of the target sequence in the gene for each primer is shown in the Appendix, Section

8.2.

DNA amplification

Prior to each reaction, all reagents, excluding primers, and all sequencing apparatus including PCR plates, lids, tips and pipettes were sterilised by 45 minutes of 254 nm wavelength UV treatment, and then PCR reactions were prepared in the sterilised UV hood (ASTEC Microflow, UK). Reaction master mixes were prepared according to the protocol in Tables 4.7 an 4.8, and aliquoted into a 96-well plate (AB gene, UK). Reagent concentrations and manufacturer's details are shown in Table 4.3. To confirm that PCR master mix was not contaminated, additional blank (master mix only) reactions were always performed. Plates were sealed with adhesive PCR film (Thermo Scientific, UK). Prepared plates were immediately transferred to a Peltier Thermal Cycler (PTC-225, MJ Research, USA), using the annealing temperatures indicated in Table 4.7 and Table 4.8, and according to the thermal cycling protocol in Table 4.4.

Reaction efficiency was tested by agarose gel electrophoresis of PCR products (see Section 4.7). Successful amplification was identified by the presence of a single strong band of appropriate length (see Figure 4.3).

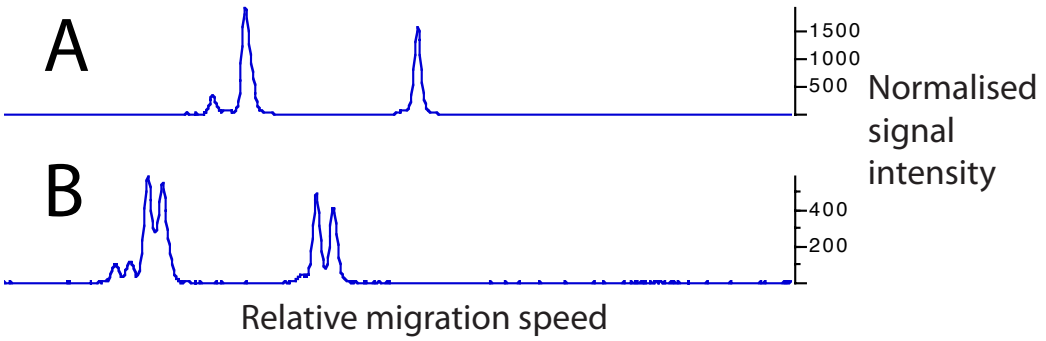


Figure 4.1: SSCP result analysis in Genotyper. The figure shows the output from Genotyper of a typical SSCP result. Lanes A and B are SSCP scan results for APC MCR 11 wild-type and tumour DNA respectively. A clear difference is observed between the scan results in lanes A and B (extra peaks are observed in lane B), indicating the presence of a mutation in this region. The position of a peak on the x-axis indicates the speed, relative to a size standard, at which the sample moved through the capillary system, and the y-axis indicates the signal intensity, normalised against the size standard.

Figure 4.2: Schematic of nested primer construction. First round primers (coloured green) were designed to amplify a region containing both the sequence of interest (coloured red) and the target sequence for the second round primers (coloured blue).

Table 4.5: APC mutation cluster region nested sequencing primer details (MCRs 1-6).

	First Round	**Second Round**
Primer name	**Sequence (5' to 3')**	**Sequence (5' to 3')**
MCR 1F	GGACAAAGCAGTAAAACCGAAC	CAAGCAGTGAGAATACGTCCA
MCR 1R	AACTACATCTTGAAAAACATATTGGA	TTTCTTGGTTAATAGAAGAAACTTTGC
MCR 2F	AAGTGGTCAGCCTCAAAAGG	GCCACTTGCAAAGTTTCTTCT
MCR 2R	GCTATTTGCAGGGTATTAGCA	TGCTTCCTGTGTCGTCTGA
MCR 3F	GATACTCCAATATGTTTTTCAAGATG	CAGACGACACAGGAAGCAGA
MCR 3R	GCCTGGCTGATTCTGAAGAT	TGGAACTTCGCTCACAGGAT
MCR 4F	CCCTGCAAATAGCAGAAATAAAA	GAAGATCCTGTGAGCGAAGTTCC
MCR 4R	AACATGAGTGGGGTCTCCTG	CAAGTCCTCTGGGGTGAGTA
MCR 5F	CAGACTGCAGGGTTCTAGTTTATC	AGAATCAGCCAGGCACAAAG
MCR 5R	CATTCCACTGCATGGTTCAC	GCAATCGAACGACTCTCAAA
MCR 6F	CCAAAAGTGGTGCTCAGACA	CACTATGTTCAGGAGACCCCA
MCR 6R	CATGGTTTGTCCAGGGCTAT	TGGAAGATCACTGGGGCTTA

Table 4.6: APC mutation cluster region nested sequencing primer details (MCRs 7-12).

	First Round	**Second Round**
Primer name	**Sequence (5' to 3')**	**Sequence (5' to 3')**
MCR 7F	TTTGAGAGTCGTTCGATTGC	GTGAACCATGCAGTGGAATG
MCR 7R	TCTCTTTTCAGCAGTAGGTGCTT	ACTTCTCGCTTGGTTTGAGC
MCR 8F	CATGCAGTGGAATGGTAAGT	GCTTAGGTCCACTCTCTCTCTT
MCR 8R	GCAGCATTTACTGCAGCTT	GGCATTATAAGCCCCAGT
MCR 9F	CAAGCGAGAAGTACCTAAAAA	CCTAAAAATAAAGCACCTACTGCTG
MCR 9R	TTCTGTATAAATGGCTCATCG	CACTCAGGCTGGATGAACAA
MCR 10F	GGTTCTTCCAGATGCTGATA	ACATTTTGCCACGGAAAGTA
MCR 10R	CTTGGTTTTCATTTGATTCTTT	GGCTGCTCTGATTCTGTTTC
MCR 11F	AATTAAGAATAATGCCTCCAGT	CAGGAAAATGACAATGGGAAT
MCR 11R	TTTACGTGATGACTTTGTTGG	TGGCATGGCAGAAATAATACA
MCR 12F	CCAAGAGAAAGAGGCAGAAAAA	GGACCTATTAGATGATTCAGATGATG
MCR 12R	TGATGGTAGAAGTTTGTACACAGG	ACTTGGTTTCCTTGCCACAG

Table 4.7: APC mutation cluster region nested sequencing PCR reagent quantities and annealing temperatures (First round). All volumes in μl and temperatures in °C. Volumes listed are for a single 25μl reaction.

	MCR											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Forward primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Reverse primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Buffer	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
dNTP	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Mg	1	3	1	1	2	2	1	2	2	2	2	1
H_20	18.25	11.25	18.25	18.25	17.25	17.25	18.25	17.25	17.25	17.25	17.25	13.25
Q		5										5
Taq	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
DNA	2	2	2	2	2	2	2	2	2	2	2	2
Annealing temp	55	60	55	55	60	60	60	55	55	55	55	55

Table 4.8: APC mutation cluster region nested sequencing PCR reagent quantities and annealing temperatures (Second round). All volumes in μl and temperatures in °C. Volumes listed are for a single 25μl reaction.

	MCR											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Forward primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Reverse primer	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
dNTP	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Mg	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Buffer	1	2	1	1	1	1	1	1	3	2	2	2
H_20	18.25	12.25	18.25	18.25	18.25	18.25	18.25	18.25	16.25	17.25	17.25	12.25
Q		5										5
Taq	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
DNA	2	2	2	2	2	2	2	2	2	2	2	2
Annealing temp	55	60	55	55	60	60	60	55	55	55	55	55

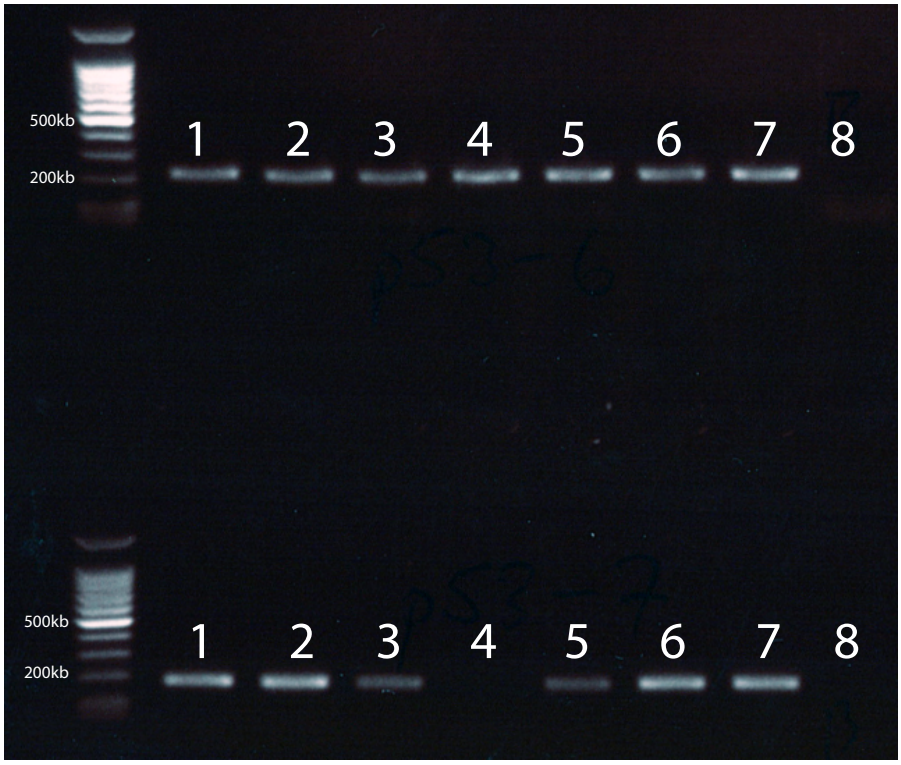


Figure 4.3: Electrophoresis gel showing successful PCR amplification. In the top row, lanes 1-7 contain successfully amplified products, identified by the presence of a bright band at around 200bps. Lane 8 has no band, corresponding to no PCR product. Lane 8 contained the blank (no template DNA) control PCR product. In the bottom row, the PCR failed for the sample in lane 4.

Sequencing

APC MCR reaction products were purified, sequenced, and analysed as described in Section 4.8.

4.6 K-ras mutation screening

Oncogenic K-ras mutations tend to occur at codons 12 and 13, and occasionally codon 61 (ANDREYEV *et al.*, 1998; WELLCOME TRUST SANGER INSTITUTE, ????). Therefore, K-ras mutation screening of micro-dissected material could be performed using only a single set of nested PCRs. Nested K-ras primers are shown in Table 4.9. Reactions were prepared according to the reagent volumes in Table 4.10. Reactions

were performed on a Peltier Thermal Cycler (PTC-225, MJ Research, USA), using the annealing temperatures indicated in Table 4.10, and according to the thermal cycling protocol in Table 4.4. Reaction efficiency was tested by agarose gel electrophoresis of PCR products. After successful amplification, reaction products were purified, sequenced, and analysed as described in Section 4.8.

Table 4.9: K-ras nested primer details.

Primer name	First Round Sequence (5' to 3')	Second Round Sequence (5' to 3')
K-ras forward	GAGTTTGTATTAAAAGGTACTGGTGGA	TTTGATAGTGTATTAACCTTAT
K-ras reverse	ATCAAAGAATGGTCCTGCAC	TATTAAAACAAGATTTACCTC

Table 4.10: K-ras nested sequencing PCR reagent quantities and annealing temperatures. All volumes in μl and temperatures in °C. Volumes listed are for a single 25μl reaction.

	First Round	Second Round
Forward primer	0.25	0.25
Reverse primer	0.25	0.25
dNTP	0.5	0.5
buffer	2.5	2.5
Mg	2	2
H_20	12.3	12.3
Q	5	5
Taq	0.2	0.2
DNA	2	2
Annealing temp	60	55

4.7 Agarose Gel Electrophoresis

Agarose gel electrophoresis separates DNA molecules according to their size. The migration rate of DNA molecules through the gel during electrophoresis depends upon the molecule size and the density of the gel. 2% agarose gels were used as they resulted in a resloving power of about 100bp in reasonable time. The addition of ethidum bromide to the gel, which intercalates between DNA base pairs and flouresces under 302nm UV light, allows the location of DNA in the gel to be visualised.

Gels were prepared by boiling an appropriate mass of agarose power in 1X TBE buffer (CRUK central services), and allowing the mixture to cool to 60°C. Ethidium bromide (Sigma, UK) was added to a final concentration of 0.25μg/ml. The gel solution was poured into a casting tray (Biorad, UK) and a comb added. The gel was allowed to set, then transferred to an electrophoresis chamber (Biorad, UK) and covered with 1X TBE

buffer. 5μl of PCR products with mixed with 2μl of Orange G loading dye (Sigma, UK) and loaded into the gel. 10μl of 1kbp ladder (Invitrogen, UK) was also loaded to provide a size reference. The chamber was connected to a DC current device (Biorad, UK) and run for 20 minutes at 120V, 400mA. DNA bands were identified using a 302 nm UV transilluminator (UVP, UK). An example of an electrophoresis gel showing successful sequencing is shown in Figure 4.3.

4.8 Sequencing

4.8.1 Purification of reaction products

Successfully amplified reaction products were first treated with EXoSAP-IT(USB, Germany). EXoSAP-IT contains exonuclease I, which removes unincorporated primers, and Shrimp Alkaline Phosphatase which removes leftover dNTPs. 2μl of EXOSAP was added to 5μl of reaction product, and then placed on a Peltier Thermal Cycler (PTC-225, MJ Research, USA), using the protocol described in Table 4.11.

Table 4.11: EXOSAP thermal cycling programme

Step	Description
1	37°C for 15 minutes
2	80°C for 15 minutes
3	15°C forever (holding temperature)

4.8.2 Sequencing reaction

Automated sequencing that used four terminator nucleotides, labelled with four distinguishable fluorophores, was performed. EXoSAP-IT products were diluted with 20μl of distilled H_20. The sequencing reaction was prepared according to the reagent quantities in Table 4.12, using ABI Biosystems, BigDye Terminator (Applied Biosystems, USA). The reaction was carried on a Peltier Thermal Cycler (PTC-225,

MJ Research, USA) according to the protocol in Table 4.13. Unincorporated dye-terminators were removed using the DyeEx 2.0 Spin Kit (Qiagen, UK), according to manufacturer's instructions.

Table 4.12: Sequencing reaction reagent quantities and annealing temperatures. All volumes in μl. Volumes listed are for a single 20μl reaction.

Reagent	Volume
Primer	1
BDT mix	10
H_20	5
Diluted EXOSAP-IT product	4

Table 4.13: Sequencing reaction thermal cycling programme.

Step	Description
1	96°C for 5 minutes
2	96°C for 30 seconds
3	50°C for 45 seconds
4	60°C for 4 minutes
5	Repeat steps 2-4 34 more times
6	8°C forever (holding temperature)

4.8.3 Sequencing and analysis

Purified sequencing products were vacuum dried using a Eppendorf 5301 Concentrator (Eppendorf, UK) for 30 minutes at 45°C. Re-suspended residues were transferred to a 48 capillary ABI 3730 DNA Analyser (Applied Biosystems, USA) for sequencing. Sequence data were analysed with the 4Peaks software (Mekentosj, Netherlands).

Chapter 5

Microdissection of Tumour Pathways in Sporadic Colon Adenomas

5.1 Aims

- Examine the development of sporadic colorectal adenomas. Determine their clonal structure.
- Infer the relative order and timing of APC and K-ras mutations in the development of the adenoma.
- Examine any relationship between the type of APC mutation and the type of K-ras mutation present in early adenomas.

5.2 Chapter Summary

The classical view of tumorigenesis in non-MMR deficient colorectal cancers begins with clonal expansion following the inactivation of the APC tumour suppressor gene in a single cell. The next genetic change in the progression towards cancer is an activating mutation of the K-ras oncogene (FEARON and VOGELSTEIN, 1990; KINZLER and VOGELSTEIN, 1997). This model of tumour development is widely accepted. However, there is evidence that K-ras mutations may precede APC inactivation in some adenomas (reviewed in PRETLOW and PRETLOW (2005)). Additionally, the two "hits" required to inactivate APC have been shown to be non-random, in that the

location of the second hit is usually determined by the first (LAMLUM *et al.*, 1999). Similar relationships between types of APC and K-ras mutations may also exist, but these have not been previously investigated.

In this chapter, the sequence of genetic events driving tumour progression in human small sporadic adenomas was examined. The aim of this work was to determine the mutational burden present in each crypt of an adenoma, and then to use these data to infer the phylogenetic relations between clones in the adenoma, and so characterise the development of the adenoma. Using laser capture micro-dissection of individual crypts, followed by direct sequencing, the mutation status of APC and K-ras genes in each crypt was determined. This information was combined to produce clonal maps of each adenoma, which showed the spread of the mutated clones through the neoplasm.

The frequency of K-ras mutation in these early adenomas was high, with $(8/12)$ neoplasms containing at least one crypt with a K-ras mutation. Further, many $(5/12)$ of the adenomas were composed entirely of crypts harbouring both K-ras and APC mutations, suggesting that K-ras mutations could have preceded APC inactivation in these adenomas. To test this hypothesis, a simple mathematical model was developed, which predicted the selective advantage required by an APC$^{-/-}$K-ras$^{+/-}$ crypts to displace APC$^{-/-}$ crypts from a small polyp. The model suggested that if K-ras mutation occurred after APC inactivation, then K-ras activation would have to bestow a large selective advantage on a clone in order for the clone to dominate the adenoma in the presumably short time that the adenoma had been growing. Hence there appeared to be two groups of adenomas, defined by whether K-ras mutation preceded or followed APC inactivation.

Next, the relationship between APC and K-ras mutations in these adenomas was examined. Adenomas were classified according to the estimated "severity" of their APC mutant genotype. No association between the inferred order of mutations and the APC mutation type was seen. The data showed a marginal excess of adenomas with a less severe APC mutation and K-ras mutations, although the excess was not statistically significant, possibly as a result of the small sample size. To investigate this hypothesis further, a much larger statistical analysis of published material on the frequency of somatic mutation of K-ras and APC in colorectal cancers was performed.

In this larger data set, the frequency of K-ras mutations showed a borderline association with the supposed milder APC phenotype. Further, the distribution of K-ras mutation types differed significantly between groups of adenomas defined by the presence of a severe or non-severe APC mutation. APC-severe cancers showed a paucity of K-ras G12D and G13D mutations compared to APC-mild cancers, whereas G12V mutations were found at approximately the same frequency in each group.

5.3 Introduction

The development of colon cancer is relatively well characterised. By collating knowledge gained from the study of colorectal neoplasms from patients with hereditary disposition to colon cancer, a number of papers by Vogelstein have suggested the order of mutations that lead from healthy tissue to a colon cancer (FEARON and VOGELSTEIN, 1990; KINZLER and VOGELSTEIN, 1997). Tumour growth is initiated by inactivation of the APC tumour suppressor gene, a *gatekeeper* mutation (KINZLER and VOGELSTEIN, 1997), and the subsequent selectively advantageous mutation tends to activate the K-ras oncogene. Further progression is through mutations of genes including p53 and SMAD4 or DCC. This pathway to cancer, shown in Figure 5.1, is parochially termed the "Vogelgram", and has become widely accepted as the baseline model of colorectal carcinogenesis.

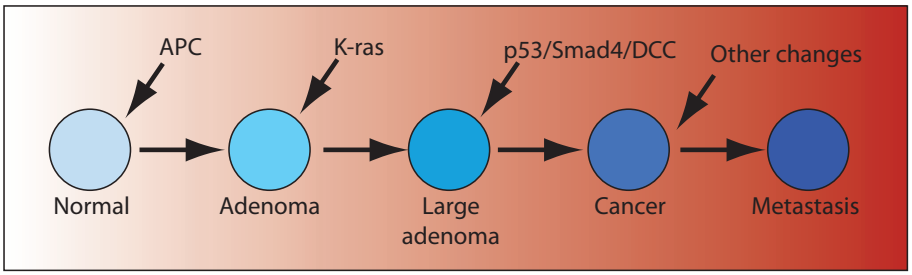


Figure 5.1: The basal model of the order genetic events in sporadic colorectal tumorigenesis. The figure depicts the widely accepted sequence of genetic alterations that transform normal colonic epithelial cells into malignant carcinoma. Figure adapted from (JONES *et al.*, 2008b).

However, there is increasing evidence that in some adenomas, K-ras mutations may precede APC mutations in tumorigenesis. APC is a large gene, and thus

detecting mutations in APC is technically difficult. In comparison, K-ras mutation is straightforward to detect, as nearly all oncogenic mutations occur at codon 12 or 13. Dysplastic aberrant crypt foci (ACF) are the putative precursors of colorectal adenomas (PRETLOW and PRETLOW, 2005). Takayama *et al.* found K-ras codon 12 mutations in 63% $(17/27)$ of dysplastic mono-cryptal adenomas from non-FAP patients, whereas APC and $\beta-$catenin mutations were never detected in these lesions (TAKAYAMA *et al.*, 2001). The prevalence of K-ras mutations in hyperplastic lesions is also very high, although whether these non-dysplastic lesions are precursors of dysplastic neoplasms remains unclear. In hyperplastic ACFs from sporadic colon cancer patients OTORI *et al.* (1998) found that 50/56 hyperplastic ACF showed K-ras codon 12 mutations, whereas none of these 56 samples had detected APC mutations. Similar findings where observed by JEN *et al.* (1994) K-ras mutations, but not APC mutations, were seen in all $(19/19)$ non-dysplastic ACFs examined. Another study found K-ras mutations in 67% $(8/12)$ of non-dysplastic ACFs from non-FAP patients (NUCCI *et al.*, 1997). SMITH (1994) did not specify the pathology of the ACFs they examined, but they found K-ras mutations in 25% (13/53) . Further, in normal mucosa, K-ras mutations were detected in approximately 10% of samples (41/390) in one study (YAMADA *et al.*, 2005), although a second study found the prevelence to be only 1% (1/100) (SMITH, 1994). By macro-dissection of normal appearing tissue located adjacent to a K-ras$^{-/+}$ tumour, ZHU *et al.* (1997) found K-ras mutations in 54% of samples $(7/13)$.

In dysplastic lesions (adenomas), the frequency of APC mutations appears considerably higher than in ACFs, consistent with the role of APC as the gatekeeper of adenomatous growth (LAMLUM *et al.*, 2000; POWELL *et al.*, 1992). K-ras mutations are still frequently found in dysplastic adenomas. Takayama *et al.* found K-ras codon 12 mutations 65% (20/31) of adenomas and the frequency of APC mutations was 78% (TAKAYAMA *et al.*, 2001). OTORI *et al.* (1998) observed lower mutation frequencies of both genes; 40% of adenomas (4/10) had K-ras mutations, and 40% had APC mutations, although only a single adenoma had mutations in both genes. JEN *et al.* (1994) observed a lower frequency, 25% (3/12), of K-ras mutations in dysplastic polyps, and a higher frequency, 83%, of APC mutations . Another study reported high proportion of K-ras mutations in sporadic adenomas (65%), although the frequency of

K-ras mutations showed a borderline association with adenoma size, and was highest in the smallest adenomas examined (MCLELLAN *et al.*, 1993).

The studies presented above suggest that the frequency of K-ras mutations is high in small dysplastic ACFs (63%), but is perhaps slightly reduced in small adenomas (range 25%-65%). In carcinomas, the frequency appears to be reduced further. The results of a very large meta-analysis found the frequency of K-ras mutations in colorectal cancers to be 37.7% ($n = 2721$) (ANDREYEV *et al.*, 1998).

In FAP patients, K-ras mutations in adenomas are observed less frequently than in sporadic colon cancer patients, whereas the incidence of somatic APC mutations is much higher. In the 8 dysplastic ACF from FAP patients examined by TAKAYAMA *et al.* (2001) K-ras mutations were found in only 13% $(1/8)$ of dysplastic ACF whereas somatic APC mutations were found in all ACF. NUCCI *et al.* (1997) observed the frequency of K-ras mutations in FAP patients to be 17% $(3/18)$. Similar low frequencies of K-ras mutations in FAP polyps, 7% $(5/75)$ and 4% $(2/51)$, were observed by FARR *et al.* (1988) and SMITH (1994) respectively. This suggests that FAP and non-FAP adenomas develop along different pathways, although the reason for this is unclear. Hence this work was restricted to the study of sporadic adenomas only.

Different mutations of the APC gene appear to bestow different selective effects on a cell (see Introduction 1.6.1). Combinations of APC mutations are selected for, so that the two hits required to inactivate the tumour-suppressor gene are non-random and so produce the "optimal" genotype, which presumably confers the greatest growth advantage (LAMLUM *et al.*, 1999; ROWAN *et al.*, 2000). APC mutations are sufficient for the growth of adenomas (LAMLUM *et al.*, 2000), however, certain types of APC mutation may influence the type of subsequent K-ras mutation, or vice-versa. The type of APC mutation that founds an adenoma could be thought of as determining the tumour "micro-environment", such that certain subsequent mutations are highly selectively advantageous in this environment, and so become fixed in the population of adenoma cells. Other mutations may confer no, or little, selective advantage in the tumour micro-environment, so not reach fixation, or even be lost through random drift.

In this chapter, the order and timing of APC and K-ras mutations in the development of small colorectal adenomas was investigated. Laser-capture micro-dissection was used to extract individual crypts from 12 small colorectal adenomas. In each crypt, the mutation cluster region of APC and the frequently mutated region of K-ras were then sequenced to determine their mutation status. Results were combined to produce a *clonal map* of the adenoma. The order and timing of APC and K-ras mutations in the development of the adenoma were then inferred by examining the clonal composition of the adenoma.

5.4 Methods

The clonal composition of 12 small adenomas was examined. All adenomas were selected from patients with no family history of cancer. Permission to perform molecular analysis of these cancers was granted by the NHS Oxfordshire Research Ethics Committee B. Detailed description of the laboratory methodology used to examine the adenomas can be found in Chapter 4. Briefly, a H&E stained slide was used to identify crypt pathology. Paraffin embedded tissue serially cut onto sterile slides and then micro-dissected on a crypt by crypt basis using laser capture micro-dissection. Crypts were sampled around the surface of each adenoma, and also from adjacent normal mucosa if available. DNA was extracted for each crypt by overnight digestion in Proteinase K. The amount of DNA in each crypt was small, so prior to direct sequencing, SSCP of the APC MCR was performed on macro-dissected tissue (needle dissection, see Chapter 4, Section 4.4.1) from each adenoma to identify mutated regions. Mutated regions were then sequenced directly using a nested PCR reaction. K-ras mutations almost always occur around codons 12 and 13, or less frequently at codon 61, so this region was sequenced directly using a nested PCR reaction.

Laboratory work was carried out in collaboration with Simon Leedham, Histopathology Department, Cancer Research UK London Research Institute.

5.5 Micro-dissection results

Results are displayed in Tables 5.1 - 5.3. Each adenoma has the status of APC and K-ras displayed for each of the crypts that were genotyped.

Table 5.1: Polyp mutation data (polyps 1-6).

Polyp ID		1		2		3		4		5		6	
		APC	**Kras**	**APC**	**Kras**	**APC**	**Kras**	**APC**	**Kras**	**APC**	**Kras**	**APC**	**Kras**
Crypt	**1**	wt	wt	wt	wt	1577X	G12D	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	2	wt	wt	1577fsX6	wt	1577X	G12D	1233fsX31	G13D	wt	G13D	1372fsX42	G12D
	3	wt	wt	1577fsX6	wt	1577X	G12D	1233fsX31	G13D	wt	G13D	1372fsX42	G12D
	4	wt	wt	1577fsX6	wt	1577X	G12A	1233fsX31	G13D	wt	G13D	1372fsX42	G12D
	5	wt	wt	1577fsX6	wt	1577X	G12A	1233fsX31	G13D	wt	G13D	1372fsX42	G12D
	6	wt	wt	1577fsX6	wt	1577X	G12A	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	7	1413fsX8	G12R	1577fsX6	wt	1577X	G12A	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	8	1413fsX8	G12R	1577fsX6	wt	1577X	wt	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	9	1413fsX8	G12R	failed	wt	1577X	G12A	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	10	1413fsX8	G12R	1577fsX6	wt	1577X	wt	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	11	1413fsX8	G12R	1577fsX6	wt	1577X	wt	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	12	wt	wt	1577fsX6	wt	1577X	wt	1233fsX31	G13D	wt	wt	1372fsX42	G12D
	13	wt	wt	1577fsX6	G13D	1577X	G12D	1233fsX31	G13D	wt	wt	wt	wt
	14			1577fsX6	G13D	1577X	G12D	1233fsX31	G13D			wt	wt
	15			1577fsX6	wt	1577X	G12D	1233fsX31	G13D			wt	wt
	16					1577X	G12D	1233fsX31	G13D				
	17					1577X	G12D	1233fsX31	G13D				

Table 5.2: Polyp mutation data (polyps 7-11).

Polyp ID		7		8		9		10		11	
		APC	**Kras**	**APC**	**Kras**	**APC**	**Kras**	**APC**	**Kras**	**APC**	**Kras**
Crypt	**1**	1306fsX14	G12D	wt	wt	wt	wt	1491fsX1	wt	1312X	wt
	2	1306fsX14	G12D	wt	wt	wt	wt	1491fsX1	wt	1312X	wt
	3	1306fsX14	G12D	1555fsX4	wt	-	wt	1491fsX1	wt	1312X	wt
	4	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	5	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	6	1306fsX14	G12D	wt	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	7	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	8	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	9	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	10	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt	1491fsX1	wt	1312X	wt
	11	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt			1312X	wt
	12	1306fsX14	G12D	1555fsX4	wt	Q1378X	wt			1312X	wt
	13			wt	wt	wt	wt			1312X	wt
	14			wt	wt					1312X	wt
	15									1312X	wt
	16									1312X	wt
	17									1312X	wt
	18									1312X	wt
	19									1312X	wt

5.5.1 APC mutations

In adenoma 1, a frame-shift mutation at codon 1413 was found in crypts 7-11. Crypts 1-6 were also dysplastic, but the frameshift mutation could not be detected in them. Presumably, these crypts contained cells with a different APC hit. Thus, this adenoma may be composed of two independently initiated APC clones.

No APC mutation could be detected in adenoma 5. SSCP screening of the APC MCR of this polyp was inconclusive, and direct sequencing of MCRs 3, 5 and 7 found no mutations. Adenoma 5 could still have harboured an undetected APC mutation that was either missed by the SSCP screen or which lay outside of the MCR. Alternatively, this adenoma may have been genuinely APC-wild-type, possibly developing due to β-catenin mutation.

Adenoma 8 had a number of crypts that were wild-type for APC (crypts 1, 2, 8, 13 and 14). Crypts 1, 2, 13 and 14 were adjacent normal crypts, whereas crypt 8 appeared to be an entrapped normal crypt (see Figure 5.3). Similarly, crypt 1 of adenoma 2 appeared histologically normal, and had no detected APC mutation. Crypts 1 and 29-31 of adenoma 12 also appeared wild-type and no APC mutations were found (see Figure 5.5).

Evidence of top-down spread of tissue was observed in adenoma 6 (see Figure 5.4). The dysplastic half of crypt 12 contained a one basepair deletion at nucleotide 4118 in APC, whereas the non-dysplastic region of the crypt was APC wild-type.

5.5.2 K-ras mutations

The majority of adenomas (67%, 8/12) contained crypts that had a mutated K-ras allele. In 4/8 of the adenomas, a mutated K-ras gene was observed in every crypt which was also mutant for APC (adenomas 1, 4, 6 and 7). 3/8 adenomas had K-ras mutations in a subset of the APC mutant crypts (adenomas 2, 3 and 12). Adenoma 5 had a K-ras G13D mutation in crypts 2-5, but no APC mutation was identified in this adenoma. The remaining 4/12 adenomas had no detectable K-ras mutation (adenomas 8-11).

In adenoma 2, the K-ras mutant crypts were localised to a small region of the adenoma, which suggested that the K-ras mutant crypt had been acquired relatively late in the

Table 5.3: Polyp mutation data (polyp 12).

Polyp ID		12	
		APC	**Kras**
Crypt	**1**	wt	wt
	2	wt	wt
	3	1370fsX3	wt
	4	1370fsX3	G12V
	5	1370fsX3	wt
	6	1370fsX3	wt
	7	1370fsX3	hom G12V
	8	1370fsX3	hom G12V
	9	1370fsX3	wt
	10	1370fsX3	G12V
	11	1370fsX3	failed
	12	1370fsX3	G12V
	13	1370fsX3	G12V
	14	1370fsX3	wt
	15	1370fsX3	G12V
	16	1370fsX3	wt
	17	1370fsX3	wt
	18	1370fsX3	G12V
	19	1370fsX3	G12V
	20	1370fsX3	G12V
	21	1370fsX3	G12V
	22	1370fsX3	G12V
	23	1370fsX3	G12V
	24	wt	hom G12S
	25	1370fsX3	G12V
	26	1370fsX3	G12V
	27	wt	wt
	28	1370fsX3	wt
	29	1370fsX3	G12V
	30	wt	wt
	31	wt	wt

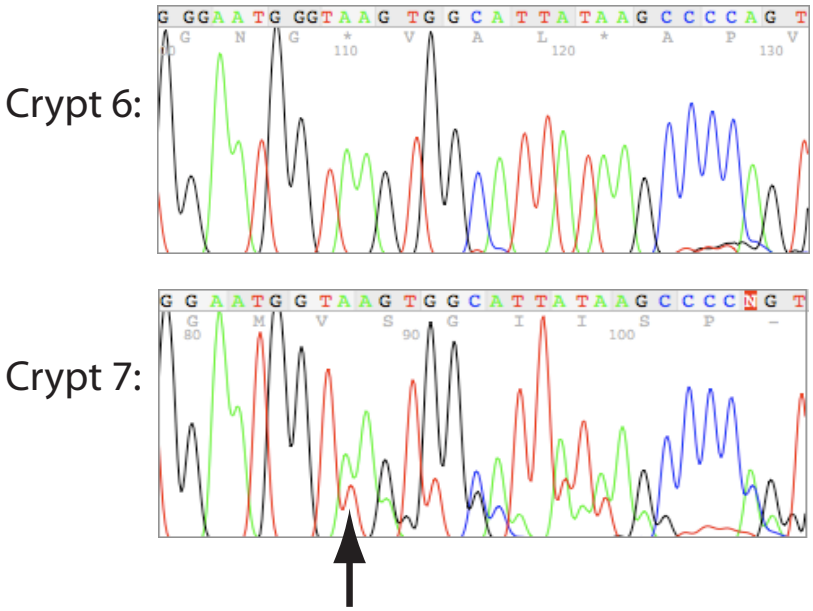


Figure 5.2: APC sequencing results from adenoma 1. Sequencing electropherogram from crypts 6 and 7 of adenoma 1. A frameshift mutation, an insertion of single T, is clearly visible in crypt 7 (indicated by an arrow), but this mutation was not present in the dsyplastic crypts 1-6, suggesting the polyp had polyclonal origins.

development of the adenoma. In adenoma 3, every crypt had identical APC mutations (mis-sense mutation at amino acid 1577), but multiple K-ras mutations were observed: crypts 1-3 and 13-17 had K-ras G12D transitions, crypts 4-7 and 9 had G12A transitions and crypt 8 and 10-12 were K-ras wild-type. Most crypts in adenoma 12 were heterozygous for G12V mutations, some were K-ras wild type, and in two crypts homozygous G12V mutations were observed (see Figure 5.6 and Figure 5.5).

Clonal maps of adenomas 8 and 12 are presented in Figures 5.3 and 5.5 respectively. In adenoma 12, the extent of multiple different mutant clones can be clearly seen. Adenoma 8 shows how polyps can contain morphologically normal crypts.

5.6 Interpretation of results

5.6.1 Timing of K-ras inactivation

This high mutation frequency (67%) compares favourably to previously published data (TAKAYAMA *et al.*, 2001; OTORI *et al.*, 1998; MCLELLAN *et al.*, 1993). K-ras

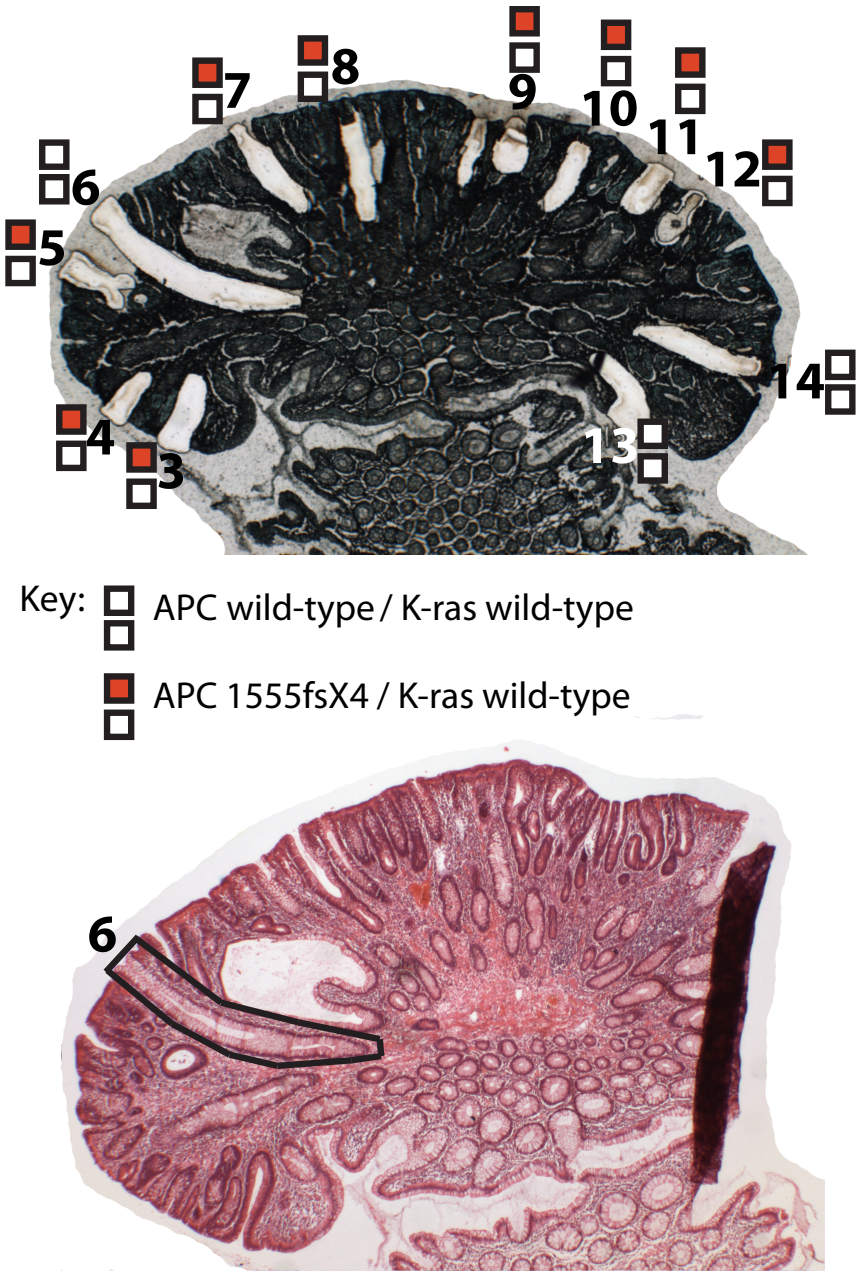


Figure 5.3: H&E and methyl green stains of adenoma 8. The top image is a micrograph of the methyl green stained laser capture slide of adenoma 8. Individual micro-dissected crypts can be clearly seen, and the mutation status of APC and K-ras is indicated by the coloured boxes. The bottom image is the H&E stained section of a contiguous section. The crypt morphology can be determined using the H&E. Crypts 1,2 13 and 14 were morphologically normal crypts. Crypts 1 and 2 were located to the left of the adenoma in non-involved mucosa. Crypt 6 appeared to be an entrapped normal crypt. The morphology of these crypts was confirmed by direct sequencing: crypts 3-5 and 7-12 carried a truncating APC mutation, a two basepair deletion at nucleotide 4665. The other crypts appeared wild-type for APC.

Figure 5.4: H&E of adenoma 6, crypt 12. Evidence of top-down spread. An APC mutation was detected in the dysplastic appearing top half of crypt 12, but not in the normal appearing basal half, suggesting that the mutated epithelial tissue was invading this crypt from the top (luminal opening) towards the crypt base.

inactivation undoubtedly occurs after APC inactivation in some adenomas: 8/12 either had no detected K-ras mutations, or contained a K-ras sub-clone. The remaining 4/12 adenomas (33%) were composed entirely of crypts that contained both APC and K-ras mutations. These adenomas could have formed in two ways:

1. APC inactivation occurred prior to K-ras mutation. APC inactivation caused the growth of the adenoma. During adenomatous growth, a crypt in the adenoma acquired a K-ras mutation. The APC$^{-/-}$K-ras$^{-/+}$ crypt had a growth advantage compared to the APC^{-} crypts. This resulted in a selective sweep of APC$^{-/-}$K-ras$^{-/+}$ crypts, which produced an adenoma composed entirely of APC$^{-/-}$K-ras$^{-/+}$ crypts.

2. K-ras mutation preceded APC inactivation. The K-ras$^{-/+}$ crypt subsequently acquired further mutations which caused the inactivation of APC and the initiation of adenomatous growth. Therefore, all crypts in the adenoma were APC$^{-/-}$K-ras$^{-/+}$.

Figure 5.5: Clonal map of adenoma 12. Adenoma 12 contained multiple different mutations. The crypt map shows their spatial arrangement in the adenomatous tissue. The majority of the adenoma is composed of clones containing both K-ras and APC mutations. The presence of multiple clones in the adenoma is clearly illustrated by the presence of two small regions that contain either crypts with only an APC mutation, or crypts with an unusual homozygous K-ras mutation.

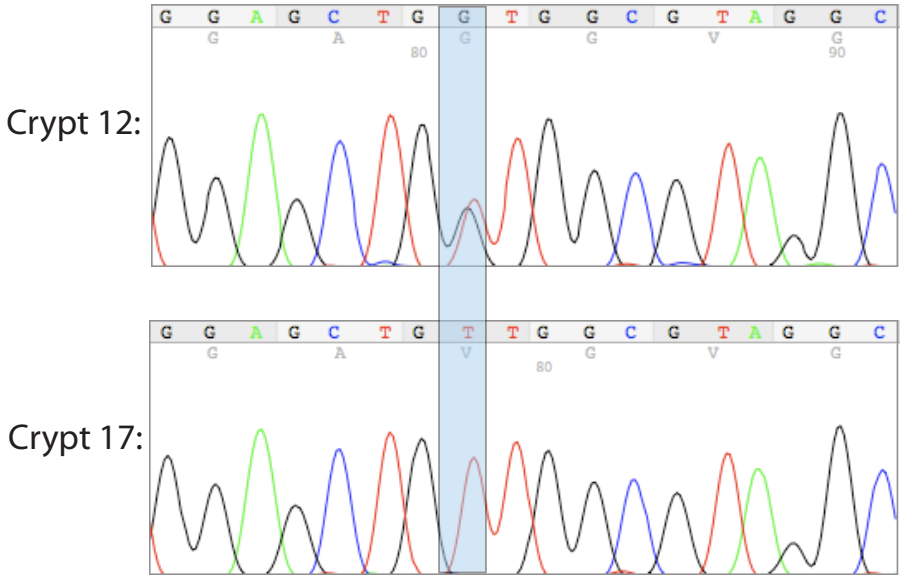


Figure 5.6: K-ras sequencing results from adenoma 12. Most crypts were heterozygous for K-ras G12V mutations (top trace), but two crypts (17 and 20) were unusual homozygous G12V mutants (bottom trace). The wild-type sequence of the codon is GGT.

If K-ras mutation occurred after APC inactivation, to reach fixation K-ras$^{-/+}$ crypts would have to rapidly sweep through the adenoma population in order to reach fixation in the presumably short period of time that the adenoma had been growing. Using this idea, a simple mathematical model of clonal competition between crypts in an adenoma was considered, which was used to estimate the selective advantage required by a K-ras$-/+$ clone to rapidly reach fixation. If K-ras mutations needed to cause a large selective advantage in order to reach fixation, it would seem more likely that the K-ras mutation occurred prior to APC inactivation. This analysis assumes that no tumour growth occurs prior to inactivation of both alleles of APC. The models are considered in section 5.6.2 below.

If K-ras mutations do precede APC mutations in some adenomas, they do not appear to be sufficient to initiate tumour growth. If oncogenic K-ras mutations were sufficient for adenomatous growth, adenomas composed entirely of K-ras mutant crypts, but with only some crypts also harbouring an APC mutation, would be expected. No such adenomas were seen. No APC mutation could be detected in adenoma 5, either as a result of a detection failure or because the adenoma was indeed APC wild-type.

Adenoma 5 contained K-ras mutations in 4/13 crypts, suggesting that even in the possible absence of an APC mutation, an oncogenic K-ras mutation did not initiate tumorigenesis. Further, if APC-deficient crypts grew out of a patch of K-ras mutant crypts, occasional K-ras$^{-/+}$APC-wild-type crypts at the margins of adenomas are expected. Again, no such crypts were seen.

5.6.2 Mathematical models of clonal competition in an adenoma

To investigate the plausibility that APC inactivation precedes K-ras mutation, simple mathematical models of clonal competition between crypts in an adenoma were developed.

Adenomas were assumed to begin growing after inactivation of APC at time $t = 0$ years. The first APC$^{-/-}$ crypt also to acquire a K-ras mutation formed at time T. Let $A(t)$ denote the number of APC mutant crypts, and $K(t)$ the number of K-ras mutant crypts.

Bottom-up spread only

The simplest model of adenoma growth assumes that tumour growth is solely via bottom-up spread (PRESTON *et al.*, 2003), so that all clonal expansion is due to crypt fission. Top-down spread, which results in one mutant crypt invading another, is neglected. Let λ_A be the net expansion rate per year of APC$^{-/-}$K-ras$^{+/+}$ crypts (net expansion rate = fission rate - extinction rate), and λ_K be the net growth rate of APC$^{-/-}$K-ras$^{-/+}$ crypts. Then the model was defined by:

$$\begin{aligned} \frac{dA}{dt} &= \lambda_A A(t); t \geq 0 \\ \frac{dK}{dt} &= \lambda_K K(t); t \geq T \\ A(0) &= 1 \\ K(T) &= 1 \end{aligned} \tag{5.1}$$

which has the solutions:

$$A(t) = e^{\lambda_A t}, t \geq 0; K(t) = e^{\lambda_K (t-T)}, t \geq T. \tag{5.2}$$

Next it was supposed that the specificity of K-ras mutant detection in an adenoma was

such that if K-ras mutations appeared in 100% of crypts examined, then at least 95% of the crypts in the adenoma contained K-ras mutant cells. Also, it was assumed that an adenoma had been growing for $\tau = 5$ years before biopsy. The growth rate of APC deficient crypts was assumed to be $\lambda_A = 0.7$ per crypt per year (see Chapter 3). Then growth rate of APC$^{-/-}$K-ras$^{-/+}$ crypts (λ_K) required to produce an adenoma that appeared to be completely composed of crypts with a K-ras mutation, given that the first K-ras mutant crypt appeared at $t = T$ years, was calculated by solving:

$$\frac{K(t)}{K(t) + A(t)} \geq 0.95 \tag{5.3}$$

for λ_K. Hence, the relative increase in growth rate required by a K-ras mutant crypt is:

$$\alpha = \frac{\lambda_K}{\lambda_A} \geq \frac{1}{\lambda_A} \frac{5\lambda_A - \log(\frac{5}{95})}{5 - T} \tag{5.4}$$

If a K-ras mutant crypt is formed shortly after the initiation of adenoma growth, the K-ras mutant crypt must divide around twice as fast as its APC mutant rivals in order to dominate the adenoma after 5 years (see Figure 5.7). Alternatively, if the K-ras crypts forms after 2.5 years of adenomatous growth, it requires a 3.7 fold increase in growth rate, whereas if it forms after four years a 9.2 fold increase is required.

However, it should be noted that if adenomas had been growing for longer than five years, then K-ras clones which occurred early in adenomatogeneous, required a smaller growth advantage to reach fixation. If equation (5.4) is considered in the limit $\tau \to \infty$, with $T \ll \tau$, then $\alpha \to 1$, indicating that K-ras mutant clones require almost no growth advantage to reach eventual fixation within the adenoma.

Top-down spread

The basal model presented above was modified to include top-down spread. The term $\gamma A(t)K(t)$ described the rate at which K-ras/APC mutant crypts invaded APC mutant crypts, where γ was the rate of top-down spread of K-ras$^{-/+}$APC$^{-/-}$ crypts per crypt per year. The crypt population in the adenoma was assumed to be well mixed so that the spatial arrangement of crypts in the adenoma could be neglected. This resulted in

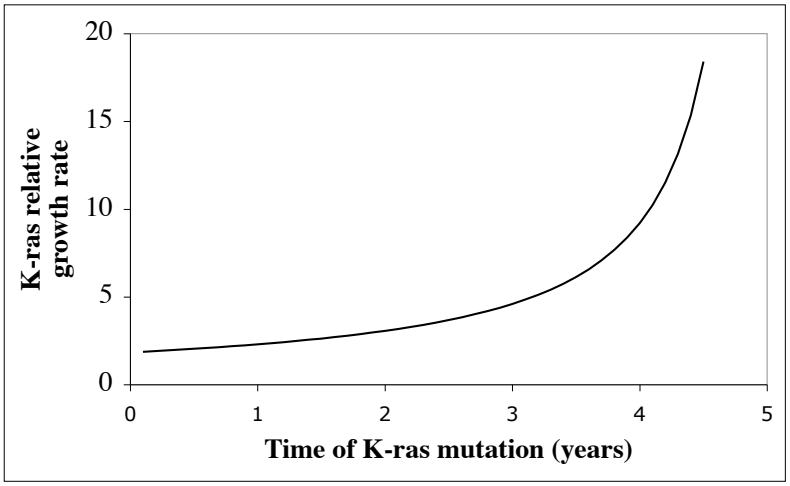


Figure 5.7: Increase in growth rate of K-ras$^{-/+}$ crypts when only bottom-up spread is permitted. Time of K-ras mutation against required relative increase in growth rate of crypts with both K-ras and APC mutations, compared to crypts with only an APC mutation, in order for the K-ras mutant crypts to dominate the adenoma after five years of growth. Only bottom-up spread (crypt fission) is permitted. If oncogenic K-ras mutation occurs late, K-ras crypts require a much higher growth rate than APC crypts to dominate the adenoma.

the equations:

$$\begin{aligned} \frac{dA}{dt} &= \lambda_A A(t), T > t \geq 0; \\ \frac{dA}{dt} &= \lambda_A A(t) - \gamma A(t)K(t), t \geq T; \\ \frac{dK}{dt} &= \lambda_K K(t) + \gamma A(t)K(t), t \geq T; \\ A(0) &= 1; \\ K(T) &= 1. \end{aligned} \tag{5.5}$$

These equations have no analytic solution, but were easily solved numerically using a Runge-Kutta type integration scheme in Mathematica (WOLFRAM RESEARCH, INC., 2005). To describe the behaviour of these equations, it was assumed that $\lambda_A = 0.7$ as described above, and that an adenoma had been growing for 5 years before biopsy. Then given that the first K-ras deficient crypt formed after T years, for each set of

parameters $\{\lambda_K, \gamma\}$ the equations 5.1 were numerically integrated, and the set of values of λ_K and γ which resulted in an adenoma composed of at least 95% K-ras mutant crypts at biopsy (see equation 5.3) was recorded.

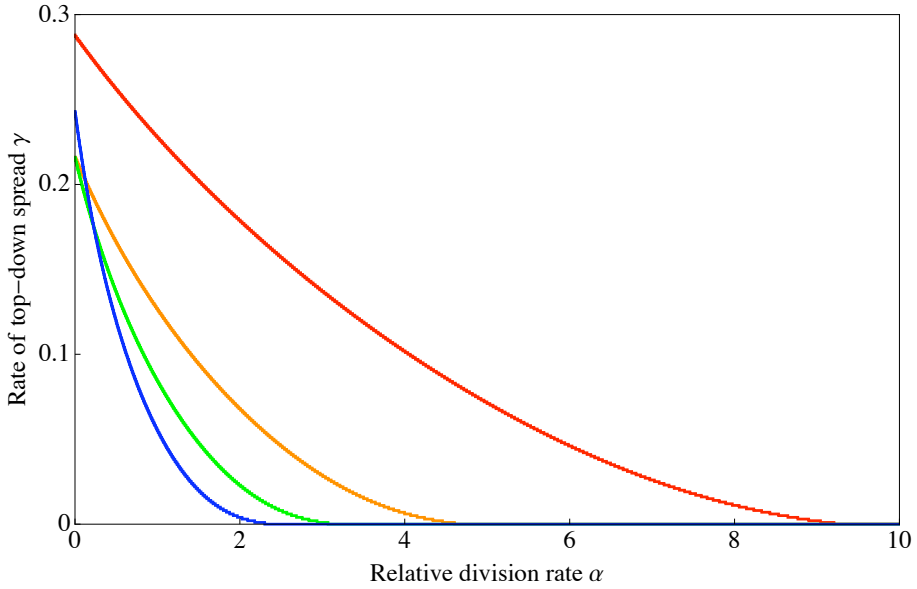


Figure 5.8: Increase in growth rate of K-ras$^{-/+}$ crypts when top down and bottom-up spread is permitted. Relative increase in growth rate of K-ras$^{-/+}$APC$^{-/-}$ mutant crypts in an adenoma initiated by APC inactivation. x-axis: relative increase in rate of crypt fission of K-ras$^{-/+}$APC$^{-/-}$, per crypt. y-axis: rate of top down spread of K-ras$^{-/+}$APC$^{-/-}$ crypts, per crypt per year. Line colour indicates time of formation of K-ras$^{-/+}$ clone (T). Blue line $T = 1$ year, green line $T = 2$ years, orange line $T = 3$ years, red line $T = 4$ years.

Figure 5.8 shows the minimum values of the relative increase in fission rate (α) and the rate of top-down spread (γ) required for K-ras$^{-/+}$ crypts to reach fixation about 5 years, given that the mutant clone formed T years after the initiation of adenomatous growth. If the K-ras$^{+/-}$ clone formed after $T = 3$ years, invading crypts require a near five-fold increase in fission rate or a high rate of top down spread.

5.6.3 Different pathways of tumorigenesis

The order and types of mutations found in an adenoma may be related to the initial genetic changes in the adenoma. The first mutation that occurs in a crypt must be a stochastic event. Phenotypic effects resulting from the initial mutation, such as

alteration of the micro-environment of the crypt, or resulting intrinsic changes in mutant cells, may alter the selective advantage bestowed by subsequent mutations, and so predispose to specific genetic alterations. The data were examined to see whether the putative order of mutations was related to the type of APC mutation present in the adenoma, since the selective advantage bestowed by an APC mutation is likely to be determined by the specific mutation type (see Introduction 1.6.1).

The hypothesis tested was that the proposed milder types of APC mutations would be more likely to require a K-ras mutation to initiate adenomatous growth, whereas severe types of APC mutation would tend to be sufficient to initiate tumour growth.

APC phenotype classification

To classify the phenotypic effect of APC mutation, the selective advantage of each APC mutation was estimated. Truncating APC mutations around codon 1309 (codons 1285-1379), that result in a single remaining 20 amino acid binding repeat, produce the most severe phenotypes in mice and humans, so presumably confer an optimal selective advantage, whereas other truncations appear to provide less of a selective advantage (see Introduction 1.6.1). Further, the location of the second hit is usually determined by the location of the first hit (LAMLUM *et al.*, 1999). Mis-sense mutations around 1309 usually result in LOH as the second hit, whereas mis-sense mutations away from codon 1309 typically cause the second hit to be a truncation elsewhere in the MCR.

Accordingly, the selective advantage of an APC mutation type, or *severity* of the genotype, was classified as follows:

- **Severe** if the adenoma harboured a truncating mutation in APC between codons 1285-1379.
- **Mild** if the adenomas did not have a truncating mutation in APC between codons 1285-1379.

Available samples (small polyp size) prevented identification of biallelic APC mutations. Accordingly, this classification is a necessary simplification.

Timing of K-ras mutation

The order of APC and K-ras mutations in an adenoma was inferred by considering the clonal composition of the adenoma. Adenomas were divided into two groups:

1. Those with an initial APC mutation. This group includes all adenomas with no K-ras mutated crypts, and those adenomas with only a proportion of the crypts containing K-ras mutations.
2. Those with an "initial" K-ras mutation. This group is composed of the adenomas with K-ras mutations in every crypt.

No association between type of APC mutation and timing of K-ras mutation

The timing of K-ras mutation showed no clear association with APC mutation type (see Table 5.6.3).

		APC mutation type	
K-ras acquitision time		**mild**	**severe**
	pre-APC	2	2
	post-APC	4	3

Table 5.4: Timing of K-ras mutation compared to APC mutation type. The data shows no association ($p = 1$ Fisher's exact test).

5.6.4 Association between APC mutation type and K-ras mutations

The data were then investigated to see whether there was an association between the type of APC mutation and the presence of a K-ras mutation. The types of APC mutation present in an adenoma might be related to whether a K-ras mutation is acquired at all. For example, if the APC mutation present in the adenoma bestowed a large selective advantage, subsequent acquisition of an activating K-ras mutation may provide limited extra selective advantage, and so the K-ras mutant clone may be unlikely to reach

fixation in the adenoma population. Therefore, in adenomas with severe type APC mutations a relative deficiency of K-ras mutations would be expected. Further, certain combinations of APC and K-ras mutations may produce optimal selective advantages, and so certain pairs of mutations will often tend to become fixed in the adenoma. Other combinations may produce less, or no, selective advantage and so be prone to loss through random drift.

The types of APC mutation were classified as above (Section 5.6.3) and the frequency of K-ras mutation with each type of APC mutation was considered. The data showed a slight excess of K-ras mutations with APC-mild phenotypes (see Table 5.6.4), although the difference was not significant ($p = 1$ Fisher's Exact Test, risk ratio 1.39).

	APC mutation type	
	mild	**severe**
K-ras wt	1	2
K-ras mutant	5	3

Table 5.5: Frequency of K-ras mutations by APC mutation type. The data shows no significant association ($p = 0.55$, Fisher's Exact Test), but there is a slight excess of adenomas with both an APC-mild type mutation and a K-ras mutation.

The lack of significant association could be the result of the small sample size. To increase the sample size, an analysis of published somatic APC and K-ras mutations was performed.

5.6.5 Population data

To examine the association between the frequency and type of K-ras and APC mutations data in a larger data set, data from COSMIC, the catalogue of somatic mutations in cancer, was used (BAMFORD *et al.*, 2004; FORBES *et al.*, 2006). COSMIC is a curated database that catalogues published mutations in genes that show frequent somatic mutations in human cancers. COSMIC release 35, published 16 Jan 2008, was used. Data are classified primarily according to gene. Data for each is in a separate

file that lists all somatic mutations recorded for that gene. Information recorded in the database includes: COSMIC Sample ID (unique sample identifier), mutated amino acid using standard nomenclature (DEN DUNNEN and ANTONARAKIS, 2001), tissue type, tissue source (for example fresh frozen or cell-line), histology, publication details, and mutation zygosity where available. Files are in comma delimited format (csv) or Microsoft Excel format (xls).

Data analysis

Data analysis was performed using R (R DEVELOPMENT CORE TEAM, 2007). Lists of somatic mutations for APC and K-ras were downloaded in csv format and imported into R. Samples that had published somatic genotypes for both K-ras and APC were identified by joining the two tables using the unique identifier "COSMIC Sample ID". This left 1722 unique samples. Next, samples where the primary tissue type was not "large intestine" were discarded, leaving 519 samples. Further, cell-line samples were removed, as cell lines may have acquired additional mutations through repeated passaging, and so not be representative of the mutation spectrum in colorectal cancers. This left 479 samples.

Many samples that had been tested for APC did not have a detected mutation (75%, 347/470). The frequency of APC mutations in colorectal cancers is thought to be around 80% (MIYOSHI *et al.*, 1992). This apparent disparity between the reported APC mutation frequency and the COSMIC database may be because of the technical difficulty in identifying mutations in APC. APC is a large gene, composed of 2843 amino-acids, and many studies genotype only a small region of the gene, so that many mutations may go undetected. Hence samples where an APC mutation was not found were excluded from the analysis, leaving 132 samples. Next, samples which had a detected APC mutation, but where the mutation was not precisely known (probably due to a technique such as SSCP being used to identify mutations) were excluded. This left a total of 114 samples with adequately characterised APC mutations.

Next, samples with a non-truncating APC mutation were excluded. Such samples were relatively rare (10/114). Although mis-sense mutations in APC may have a prognostic effect (FRAYLING *et al.*, 1998), their pathogeniticity compared to truncating

mutations is unclear. After exclusion of mis-sense mutations 104 samples remained. A single cancer had two K-ras mutations listed. This was also excluded, leaving 103 samples.

Next APC mutations were classified according to their genotype. Samples which had a truncation mutation between amino acids 1285 and 1379, which resulted in retention of a single 20 amino-acid binding repeat, were labelled as "APC-severe"; otherwise the sample was described as "APC-mild". A few samples (4/104) had more than one listed somatic mutation in APC. It was assumed that an APC-severe genotype was dominant, so that if either or both of the mutations were a truncation between amino acids 1285 and 1379, the cancer was classified as APC-severe, otherwise it was classified as APC-mild.

The 103 cancers had been curated from thirteen papers (references: (YUAN *et al.*, 2001; DEUTER *et al.*, 1996; KANAOKA *et al.*, 1996; JEN *et al.*, 1994; MIYAKI *et al.*, 2003, 2002; LILLEBERG *et al.*, 2004; DOMINGO *et al.*, 2004; KIM *et al.*, 2004a; AOKI *et al.*, 1994; MIYAKI *et al.*, 1997; JEON *et al.*, 2008; DE ABAJO *et al.*, 2007)). These papers were reviewed to confirm that all cancers were sporadic. If cancers from patients with germline APC mutations were included in the analysis, this would have systematically under-represented cancers with severe-type somatic APC mutations. Cancers with a germline mutation around codon 1309 tend to have a second hit through allelic loss (LAMLUM *et al.*, 1999). COSMIC only lists somatic mutations, and cancers with identified somatic LOH events would have been excluded by the filtering described above, whereas FAP patients with a germline mutation elsewhere in APC tend to have a truncating mutation in the MCR as their second hit, so would have been included after the data were filtered.

The following analyses were performed:

1. The association between the frequency of K-ras mutations with APC-severe and APC-mild phenotypes (Fisher's Exact Test).

2. The association between the type of K-ras mutations with APC-severe and APC-mild phenotypes (Fisher's Exact Test).

Results

Table 5.6 shows the frequency of K-ras mutations with APC-severe and APC-mild phenotypes. K-ras mutations show a possible association with an APC-mild phenotype ($p = 0.055$, Fisher's Exact Test, RR of K-ras mutation with an APC-severe phenotype $= 0.52$).

Table 5.6: Frequency of K-ras mutations with APC-severe and APC-mild phenotypes.

	APC-mild	**APC-severe**
K-ras wt	36	32
K-ras mutant	26	9

Table 5.7 shows the frequency of K-ras mutation types with APC-severe and APC-mild phenotypes. K-ras G12D and G13D mutations are rarely seen with the APC-severe genotype (risk ratios 0.11 and 0.18 respectively, p-values (compared to the frequency of APC-mild and APC-severe mutations in cancers with no K-ras mutation) are $p < 0.006$ and $p < 0.07$ respectively). K-ras G12V occur with approximately the same frequency with APC-severe and APC-mild mutations (risk ratio 0.84, $p = 1$).

5.7 Discussion

Small adenomas, the precursors of colorectal cancers, have been growing for a presumably short time and therefore their clonal composition represents a snapshot of the initial events of colorectal carcinogenesis. The frequency of K-ras mutations in the adenomas examined in this chapter was high (66%) and compared favourably to other studies of early dysplastic lesions (TAKAYAMA *et al.*, 2001). APC mutations were found in all but one of the adenomas examined (92%, 11/12), which was consistent with APC inactivation as the gatekeeper of adenomatous growth (LAMLUM *et al.*, 2000; POWELL *et al.*, 1992). No evidence was observed that K-ras mutations were sufficient to initiate adenomatous growth, since no adenomas that were homogeneous for K-ras mutations, and heterogeneous for APC mutations, were seen. Contrary to the

Table 5.7: Frequency of K-ras mutation types with APC-severe and APC-mild phenotypes.

	APC-mild	**APC-severe**
K-ras wt	35	32
p.G12A	1	0
p.G12D	13	1
p.G12R	1	0
p.G12S	0	4
p.G12V	4	3
p.G13D	7	1
p.Q61H	1	0

findings of ZHU *et al.* (1997), no K-ras mutations were observed in normal appearing crypts that were situated adjacent to K-ras$^{-/+}$ adenomas. A possible explanation for the discrepancy is that, Zhu *et al.* manually macro-dissected tissue adjacent to adenomas, and then used a particularly sensitive assay to screen for K-ras mutations. Hence their detection of K-ras mutations in normal appearing epithelium could be because of small amounts of contaminating tumour tissue. The laser capture techniques used in this chapter minimise the risk of contaminating tissue.

K-ras mutations were likely to have occurred prior or close to the time of APC inactivation in some adenomas, as otherwise a late developing K-ras$^{-/+}$ clone would have required a very large selective advantage in order to reach clonal dominance in the presumably short time that the adenoma had been growing. Further, no association was found between the type of APC mutation and the inferred order of APC and K-ras mutations. Hence, it appears that if K-ras mutations do occasionally precede APC mutations in adenomatous growth, the order of mutations is likely to be the result of a random event, rather than the order of mutations representing a distinct pathway of tumorigenesis.

APC is a tumour suppressor gene, and hence requires two allelic hits for inactivation and subsequent tumour growth. In comparison, the K-ras oncogene requires only

a single hit for activation. Assuming that individual K-ras and APC hits occur with equal likelihood, which may be a gross over-simplification (since tumorigenic APC mutations can occur in a large region whereas oncogenic K-ras mutations are generally restricted to two codons), then APC inactivation should occur many orders of magnitude less frequently than oncogenic K-ras mutation. However, since K-ras mutation is insufficient to initiate adenoma growth, some K-ras mutations may go undetected, whereas presumably most cells that acquire two APC hits result in a potentially detectable adenoma (POWELL *et al.*, 1992; LAMLUM *et al.*, 2000). The high frequency of K-ras mutations observed in early adenomas may be the result of a relatively high rate of somatic K-ras mutations, that behave as "passenger" mutations, until the subsequent inactivation of APC. Alternatively, somatic K-ras mutation may initiate clonal expansion without causing overt morphological changes to the crypt - a field cancerization effect.

A relatively high rate of somatic K-ras mutation is further supported by the finding that 2/12 of the adenomas examined in this chapter contained multiple different K-ras clones. This relatively high frequency (17%) of mutiple K-ras clones is much greater than the frequency observed in colorectal cancers (0.029%)[1]. Previous studies, however, may have failed to detect minority K-ras clones, as genotyping would have been performed on macro-dissected samples that aggregated the DNA from many hundreds of crypts. Since the detected frequency of multiple K-ras clones in cancers is low, it is likely that if the two adenomas with multiple K-ras clones had continued to develop, one of the two K-ras clones would have out-grown the other and produced a K-ras monoclonal adenoma. Similarly, to reconcile the high frequency of K-ras mutations observed in early adenomas (25-66%, see Section 5.3) compared to the lower frequency observed in cancers (37.7%) (ANDREYEV *et al.*, 1998), some K-ras mutant clones in early adenomas must subsequently be lost during further tumour development, either as a result of genetic drift or out-growth of a more advantegous clone in the adenoma.

Whether a particular type of K-ras mutation will reach fixation in an adenoma may depend both on the type of the K-ras mutation and the molecular characteristics

[1] Analysis of K-ras mutations listed on COSMIC (WELLCOME TRUST SANGER INSTITUTE, ????).

of the adenoma, namely the type of founding APC mutation. The analysis of published data of the frequency of APC and K-ras mutation types in sporadic colon cancers performed in this chapter revealed a borderline paucity of K-ras mutations that were found with optimal type APC mutations (termed APC-severe), compared to the frequency of K-ras mutations found in cancers with sub-optimal APC mutations (termed APC-mild). Assuming that K-ras mutations occur at the same frequency in both groups, this data could be explained if K-ras mutations confer a greater selective advantage when coupled with a sub-optimal APC mutation. When coupled with a milder APC mutation, a K-ras mutant clone would be more likely to reach fixation, whereas when coupled with a severe APC mutation, the K-ras mutant clone may be more frequently lost through genetic drift. Further, K-ras G12D and G13D were observed significantly less often with APC-severe mutations, whereas K-ras G12V mutations were found at the same frequency with either class of APC mutation. This possibly suggests the combination of K-ras G12D or G13D mutations with APC-severe mutations does not produce a significant selective advantage, possibly because APC-severe type cells have a sufficient increase in growth rate as to render subsequent K-ras G12D or G13D mutations inconsequential. In contrast, K-ras G12V mutations were observed at the same frequency irrespective of type of APC mutation, so it appears that this mutations may always provide a significant selective advantage. Interestingly, K-ras G12V mutations are the only K-ras mutation type associated with a worse prognosis (ANDREYEV *et al.*, 2001), so presumably K-ras G12V mutations cause a more aggressive phenotype.

5.8 Conclusion

Laser capture micro-dissection, coupled with optimised direct sequencing techniques, can be used to observe the clonal structure of small adenomas. This clonal structure is indicative of the first clonal expansions that led to the formation of the adenoma.

K-ras mutations are frequently found in early adenomas, and in some cases activating K-ras mutations may occur prior to APC inactivation, although the order of mutations appears to be a random event. K-ras mutations alone appear to be insufficient to initiate adenomatous growth. K-ras mutations appear to be more frequently found in adenomas

and cancers with sub-optimal APC mutations, although the association is borderline and should be investigated further. Analysis of a public data set suggests that K-ras mutation types may show a non-random distribution between APC mutation types, with G12D and G13D mutations showing a strong association with less severe APC mutations.

5.9 Further work

5.9.1 Mouse models

Interactions between APC and K-ras mutations should be examined in a suitable mouse model. The behaviour of the murine Apc gene is very similar to the human homologue, APC. Consequently, the mouse is used as a model organism for investigating colorectal cancer development. The tumour burden in mice with the hypothesised "just-right" combinations of Apc and K-ras mutations should be compared to the tumour burden in mice with non-optimum combinations of mutations. Just-right combinations of Apc and K-ras mutations should produce the most severe phenotypes.

Of particular interest are the K-ras G12D and G12V mutations; these are the most frequently occurring K-ras mutations in humans cancers (ANDREYEV *et al.*, 1998). In this chapter, it was shown that K-ras G12D are rarely found with APC-severe type mutations, whereas K-ras G12V mutations are found at the same frequency with both APC-severe and APC-mild mutations. This suggests that an acquisition of a K-ras G12D mutation only provides a significant selective advantage in the presence of an APC-mild mutation, whereas K-ras G12V mutations always bestow a selective advantage. To test this four strains of mice should be created: Apc-mild/K-ras G12D, Apc-mild/K-ras G12V, Apc-severe/K-ras G12D and Apc-severe/K-ras G12V, and comparing the phenotypes of these mice to two groups of control mice that have either Apc-mild and Apc-severe mutations.

Severity of the cancerous phenotype in the mice could be measured by tumour burden. Molecular milestone events, such as p53 or p21 mutation, or histological features such as invasion of the muscular layer, could also be assessed.

The results presented in this chapter suggest the following hypotheses to test in the

mouse models:

1. Apc-severe/K-ras G12V mice should produce a more severe phenotype than Apc-severe/K-ras G12D mice.

2. The phenotype of Apc-severe/K-ras G12D mice will be comparable to the control Apc-severe mice.

3. Both Apc-mild/K-ras G12D and Apc-mild/K-ras G12V mice will show a more severe phenotype than the control APC-mild mice.

Additionally, phenotypes of APC-mild/K-ras G12D and APC-mild/K-ras G12V mice should be compared to Apc-severe mice. This comparison may show whether the combination of non-optimum Apc mutation and an additional K-ras mutation produces a more severe phenotype than an Apc-severe mutation alone.

Mouse types

The Min-mouse carries a germline nonsense mutation in Apc at amino acid 850 (MOSER *et al.*, 1990; SU *et al.*, 1992). In the framework presented above, this mouse will tend to produce tumours with the equivalent of an APC-mild phenotype. Recently, the Tomlinson lab have developed a new APC mutant mouse which has a germline missense mutation in Apc at codon 1322. Correspondingly, this mouse has the equivalent of an APC-severe phenotype.

K-ras G12D mutations are lethal in the mouse embryo (JOHNSON *et al.*, 1997; KOERA *et al.*, 1997). Further, somatic activation of oncogenic K-ras, through spontaneous recombination in transgenic mice, causes widespread neoplastic and developmental defects (TUVESON *et al.*, 2004), particularly in the lungs (JOHNSON *et al.*, 2001). Hence, to study the effect of K-ras mutations in the development of colorectal neoplasms, the confounding effects of oncogenic K-ras activation in other tissues must be accounted for. To achieve this, transgenic mice with inducible oncogenic K-ras should be used, which will allow for targeted expression of mutant K-ras in the bowel of adult mice. Existing transgenic mice with conditional oncogenic K-ras mutations are reviewed in (JANSSEN *et al.*, 2005), and both conditional K-ras G12D and G12V transgenic mice are available.

5.9.2 Larger colorectal cancer cohort

Insufficient data are currently available to reliably detect a weak association between the frequency of K-ras mutations and the APC-mutation type. A study with much larger data sets is required to reliably detect whether K-ras mutations are non-randomly associated with APC-mutations.

The sample size required in the APC-mild and APC-severe groups to confidently detect a difference between the frequency of K-ras mutations in each group was calculated. The required power of the study was set at 90%, and required confidence levels of 95%, 99% and 99.9% were considered. If there is no association between the frequency of K-ras mutation and APC-mutation type, the frequency of K-ras mutations was assumed to be 0.4, as observed in a large meta-analysis of the frequency of K-ras mutations in colorectal cancers (ANDREYEV *et al.*, 1998). Then, sample size was calculated for various levels of the relative risk of finding a K-ras mutation mutation with an APC-mild mutation, rather than an APC-severe mutation.

The sample size is given by:

$$n \geq \frac{\left(u\sqrt{\pi_0(1-\pi_0)+\pi_1(1-\pi_1)}+v\sqrt{2\widehat{\pi}(1-\widehat{\pi})}\right)^2}{(\pi_1-\pi_0)^2} \tag{5.6}$$

where $\widehat{\pi} = \frac{1}{2}(\pi_0 + \pi_1)$, and π_0 and π_1 are the proportions of cancers with K-ras mutations in the APC-severe and APC-mild groups respectively. u and v are percentage points of the normal distribution that correspond to $100\%-$power and $100\%-$confidence, respectively (KIRKWOOD and STERNE, 2003).

Calculated minimum required sample size is shown in Figure 5.9.2. If the relative risk of finding APC-mild and K-ras mutations, compared to APC-severe and K-ras mutations, is less than 1.57, samples sizes greater than 200 cancers are required to detect a significance difference between the groups with 95% confidence. Hence, the data presented here were insufficient to detect confidently a mild to moderate association between the frequency of K-ras mutation amongst the APC-genotype groups.

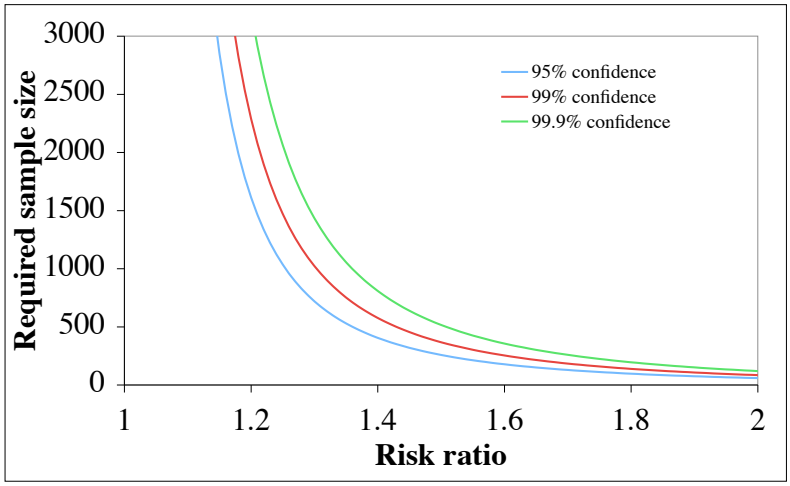


Figure 5.9: Required sample size for detecting the association between K-ras mutations and an APC-mild genotype. The relative risk is the ratio of the probabilities of observing K-ras$^{-/+}$APC-mild cancers to K-ras$^{-/+}$APC-severe cancers.

5.9.3 Clonal analysis of neoplasms

This work has optimised DNA yield and genotyping techniques for small volumes of paraffin embedded tissue. Routinely, a minimum of 6 PCR reactions could be performed on the DNA from a single crypt. These techniques could be applied to further characterise the development of colorectal adenomas, for example by combining these techniques with methylation analysis (YATABE *et al.*, 2001; KIM *et al.*, 2004c; KIM and SHIBATA, 2004; NICOLAS *et al.*, 2007) to infer expansion rates of sub-clones and better characterise the phylogenetic relations between sub-clones. The techniques could be applied to study neoplastic growth in other tissues, for example elsewhere in the gastrointestinal tract, or in the lungs.

5.9.4 Mathematical modelling of clonal competition

Mathematical modelling of the formation of APC and K-ras mutant clones, and subsequent competition between clones, would provide useful further insight into the clonal dynamics of early adenomas. In particular the expected genotypes of

cancers, that result from selective couplings of particular mutations, could be predicted. Recently, Beerenwinkel *et al.* developed a simple Wright-Fisher type model of clonal expansion, that described sub-clone dynamics during cancer growth (BEERENWINKEL *et al.*, 2007). This model could be used, but modified slightly to include mutations prior to adenomatous growth, and also by allowing the relative increase in selective advantage bestowed by a mutation be determined by the mutation type. The effect of "optimal" combinations of mutations could also be investigated in this model framework.

Chapter 6

Low Level Microsatellite Instability

This work has been published in the Journal of Pathology (GRAHAM *et al.*, 2008).

6.1 Aims

- Examine the frequency of microsatellite slippage in non-MSI-H cancers.
- Determine whether low-level microsatellite slippage can be explained by a low rate of normal somatic slippage, or if an increased slippage rate is required.

6.2 Chapter Summary

Microsatellites (MSs) are short, repetitive, sequences of DNA that occur throughout the genome. An MS locus consists of a short motif, such as CA, that is repeated a number of times. Microsatellite instability (MSI) is a common and well characterised form of genetic instability in colorectal cancers (reviewed by GRADY (2004)). Many cancers have one or more MS loci that show changes in the number of motif repeats when compared to the germline (IONOV *et al.*, 1993). Cancers which have many loci that show a high degree of variability in motif repeats tend to have germline mutations in the mismatch repair (MMR) genes (DE LA CHAPELLE, 2004; HOEIJMAKERS, 2001; KINZLER and VOGELSTEIN, 1996), or to have methylation of the MLH1 promoter, which is associated with protein loss (KANE *et al.*, 1997; VEIGL *et al.*, 1998). These (epi)genetic changes impair the DNA repair machinery of the cell, causing an increase

in the rate of MS slippage, so slippages accumulate at many loci. Consequently, MMR-deficient cancers are termed microsatellite instability - high cancers (MSI-H). It has been suggested that some cancers with a lower level of MS slippage (slippage at only a few loci), and without MMR deficiency, form another distinct pathological group (BOLAND *et al.*, 1998). These cancers are labelled low-level MS unstable (MSI-L). Analogous to MMR-deficiency in MSI-H cancers, putative MSI-L cancers are thought to have an underlying defect, possibly transient, that increases their MS slippage rate. As yet, no such defect has been reliably identified, and further, no associations of MSI-L with any clinico-pathological variables have been firmly established (TOMLINSON *et al.*, 2002).

Due to the relatively high rate of somatic MS slippage in normal tissue (BHATTACHARYYA *et al.*, 1994; HANFORD *et al.*, 1998; KRUGLYAK *et al.*, 1998; LAI and SUN, 2003) and the large number of cell divisions that occur during carcinogenesis, some MS slippage is expected in a cancer. This has led some authors to question whether the putative MSI-L group of cancers do indeed form a distinct pathological group (LAIHO *et al.*, 2002; TOMLINSON *et al.*, 2002), and instead suggest that MS slippage in these cancers is the result of normal somatic slippage during carcinogenesis.

In this chapter, data on microsatellite slippage in 47 colorectal cancers were analysed. Cancers were genotyped at $N = 9$ MS loci by Sarah Halford, Molecular and Population Genetics Laboratory, Cancer Research UK London Research Institute (HALFORD, 2005). Cancers were macro-dissected into four distinct regions that were genotyped separately. MSI-H status was assessed by the presence of slippage at the BAT-26 locus (STONE *et al.*, 2000). A mathematical model was developed, which predicted the amount of MS slippage expected in cancer due to normal somatic slippage. The predictions of the model were compared to the degree of slippage observed in the cancer cohort, to assess the evidence for an increased slippage rate during the growth of each cancer. If the model predicted that the slippage observed in a particular cancer could be explained by background slippage alone, then it was concluded that the cancer had not acquired a MSI-L phenotype. If the model predicted that the degree of slippage expected from normal somatic slippage was inadequate to explain the slippage in a cancer, then it was assumed that MS slippage rate of the cancer was increased during

tumorigenesis. Thus the model facilitated the evaluation of the *null hypothesis* that all MS slippage observed in a putative MSI-L cancer resulted from only a normal rate of somatic mutation.

5/47 cancers were confirmed as MSI-H by the presence of slippage at the BAT-26 locus. These cancers exhibited a degree of slippage that was more severe than predicted by the mathematical model. This comparison provided validation that the model was sensitive to an increase in the MS slippage rate during tumorigenesis. Of the remaining 42 cancers, 10/42 showed a level of slippage in at least one region that was too severe to be explained by the mathematical model, when the slippage rate was set to $\mu = 10^{-5}$ per locus per division. If the slippage rate was increased slightly to $\mu = 5 \times 10^{-5}$ per locus per division then none of the non-MSI-H cancers showed a degree of MS instability that was too severe to be explained by the model. This suggested that most low-level microsatellite slippage could be explained by a low rate of normal somatic slippage, and if MSI-L had occurred in these cancers, it was a relatively infrequent event.

Next it was assumed that the model, with a slippage rate of $\mu = 10^{-5}$ per locus per division, correctly identified whether a cancer had acquired an MSI-L phenotype. The sensitivity and specificity of counting the number of "unstable" loci to identify MSI-L cancers was then considered. Requiring a single unstable locus for MSI-L classification gave poor specificity, whereas requiring more than one unstable locus gave poor sensitivity. Considering the degree of instability present at each locus, rather than just the number of unstable loci, appeared to be a more reliable indicator of MSI-L status.

Associations with predicted MSI-L status and patient age at biopsy, site of cancer, and cancer stage revealed no significant associations, although this may have been due to the relatively small size of the study.

Heterogeneity in the degree of MSI observed in different regions of a cancer was assessed. Only one cancer had a degree of MSI indicative of MSI-L in all four macro-dissected regions. In other cancers, MSI that was indicative of MSI-L was restricted to localised regions. Hence if an underlying genetic change did cause MSI-L in these

cancers, it tended to occur late in tumour development.

The model of normal somatic slippage in colorectal cancers could explain most of the low-level microsatellite slippage observed in the cohort of cancers. Hence, there was little evidence for a discrete group of MSI-L cancers. If an underlying genetic change did cause MSI-L, it appeared to be a relatively uncommon event that occurred late during tumorigenesis.

6.3 Introduction

The importance of an increased mutation rate is tumorigenesis is contested. Tumorigenesis is a multi-stage process, such that the transformation of a normal cell to an invasive carcinoma requires multiple mutations (FEARON and VOGELSTEIN, 1990; HANAHAN and WEINBERG, 2000; JONES *et al.*, 2008b). Since the rate of somatic mutations is relatively low, of the order of 10^{-10} per base per division, some authors have argued that an increase in the somatic mutation rate during tumour formation is necessary for the acquisition of multiple mutations in reasonable time (RAJAGOPALAN *et al.*, 2003; KOMAROVA *et al.*, 2002; NOWAK *et al.*, 2002). Other authors have argued that successive waves of clonal expansion, which results in many cells which are susceptible to further oncogenic mutations, provides an alternative mechanism for a cancer cell lineage to acquire multiple mutations (TOMLINSON *et al.*, 1996; MOOLGAVKAR and LUEBECK, 2003; BEERENWINKEL *et al.*, 2007).

A particular type of genetic instability, called *microsatellite instability*, has been extensively studied in colorectal cancers (see reviews by DE LA CHAPELLE (2003) and GRADY (2004)). Microsatellite (MS) loci are a repetitive sequences of DNA found throughout the genome. A microsatellite locus consists of a short motif of bases, such as CA, which is repeated a number of times. The *length* of an MS locus is defined as the number of motif repeats at the locus. For example, the locus CACACACACACACACA has a length of 8. Microsatellite loci are susceptible to polymerase *slippage*, which can cause a change in the number of motif repeats (see Introduction Section 1.7.2) and hence MS slippages are observed relatively frequently compared to other mutation types.

Approximately 15% of colorectal cancers show extensive microsatellite slippage at most MS loci examined (DE LA CHAPELLE, 2003). Accordingly, these cancers are labelled as having a high level of microsatellite instability, and are termed *MSI-High* cancers (MSI-H). MSI-H cancer cells have defective mis-match repair machinery; either mutations of the MMR genes (HOEIJMAKERS, 2001; KINZLER and VOGELSTEIN, 1996; DE LA CHAPELLE, 2004) or methylation of the MLH1 promoter which is associated with protein loss (KANE *et al.*, 1997; VEIGL *et al.*, 1998) (see Introduction section 1.7.2), which results in an impaired ability to correct MS slippage during DNA replication, and so MS slippages accumulate during the many rounds of cell division in tumour growth. Since microsatellites are located throughout the genome, the increased frequency of MS slippage increases the likelihood of a mutation in a tumour suppressor or oncogene, and so drives tumorigenesis.

Some cancers show a lower level of microsatellite slippage, and have been termed *MSI-Low* cancers (MSI-L) (BOLAND *et al.*, 1998). These cancers tend to have MS slippage at only a few of the loci that were genotyped, and the degree of slippage observed at each locus is significantly less than in MSI-H cancers (see Figure 6.3). MSI-H cancers tend to have multiple extra peaks compared to the germline, whereas MSI-L often show only a change of a single motif repeat. This MSI-L group of cancers is less well defined than MSI-H cancers. Typically, putative MSI-L cancers are identified by counting the number of loci that show slippage (reviewed in TOMLINSON *et al.* (2002)). Typically, MS-stable cancers (MSS) are classified by having no loci that show slippage, MSI-L cancers classified by having between 1-3 "unstable" loci, or less than 30% of the loci examined showing instability, whereas cancers that have more loci showing slippage are labelled MSI-H. MSI-H status can also be reliably identified by slippage at particular *indicator* loci. For example, the BAT-26 MS locus, located in intron 5 of the mismatch repair gene MSH2, has been shown to be both sensitive and specific for identifying MSI-H cancers (STONE *et al.*, 2000). No such indicator loci have been identified for MSI-L cancers (LAIHO *et al.*, 2002; BOLAND *et al.*, 1998).

Previous research has examined the associations between putative MSI-L cancers and various clinical and pathological factors. To date, no significant associations have been found. One group reported an association between MSI-L and K-ras mutations (JASS

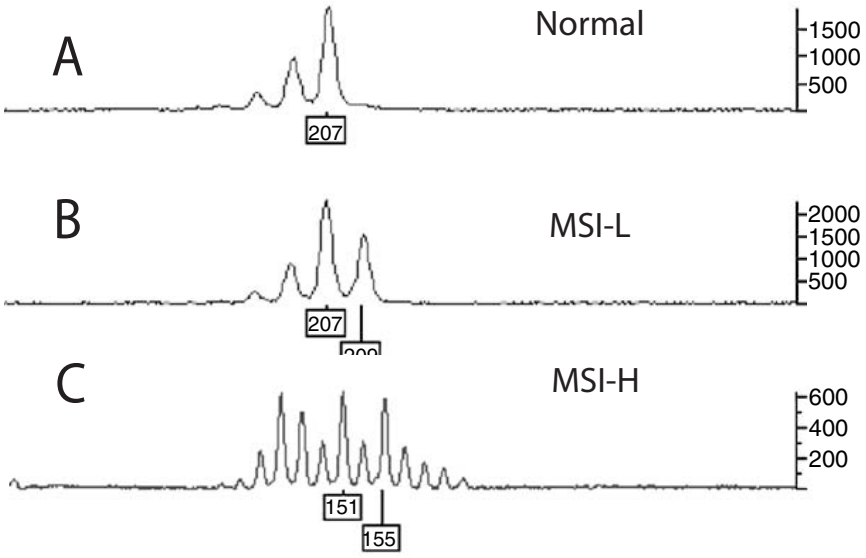


Figure 6.1: Genescan traces from MSI-H, MSI-L and MSS cancers. (A) Genescan trace of normal DNA. Small amounts of slippage are observed after PCR even in normal (wild-type) DNA, due to slippage during the DNA amplification process. (B) Typical trace from a putative MSI-L cancer. A single extra allele that differs by a single motif repeat is observed. (C) Trace from an MSI-H cancer shows severe instability: there are many extra alleles, some of which show large deviations in length compared to the germline. Image reproduced with permission from KETS *et al.* (2006).

et al., 1999), but numerous other studies have not found the association (see for example HALFORD *et al.* (2005); LAIHO *et al.* (2002) or review in TOMLINSON *et al.* (2002)). Further, the same group reported an association between MSI-L and methylation of O-6-methylguanine DNA methyltransferase (MGMT) (WHITEHALL *et al.*, 2001), but again a subsequent study did not observe such an association (HALFORD *et al.*, 2005). A relatively large study of patients with sporadic stage C colorectal cancers ($n = 255$ patients) found a borderline association between MSI-L and reduced cancer specific survival (WRIGHT *et al.*, 2005), although no stage-specific associations have been noted in other studies (see review in TOMLINSON *et al.* (2002)). Recently, a study of ($n = 98$) colorectal cancers found an association between MSI-L and a reduced immune response (reduced numbers of lymphocytes in the cancer) (KETS *et al.*, 2006). The lack of a reliable definition for MSI-L has made looking for such associations difficult.

Somatic MS slippage occurs at a relatively high rate in apparently normal cells. Estimates from cancer cells lines suggest a rate of around $10^{-7} - 10^{-5}$ per locus per cell per division in MMR-proficient cancer cell lines (BHATTACHARYYA *et al.*, 1994; HANFORD *et al.*, 1998), and analysis of slippage rates between human generations range between about 10^{-8} and 10^{-3} (LAI and SUN, 2003; KRUGLYAK *et al.*, 1998), with the mean slippage rate for a dinucleotide repeat about 2×10^{-4} (LAI and SUN, 2003). Cancer growth involves many rounds of cell division, due in part to the number of amplifying divisions required to produce the large number of cells in a cancer (around 10^9 in a cancer of around 1cm^3), and also because of repeated selective sweeps as cancer cells acquire new selectively advantageous mutations. Hence, the combination of a large number of divisions coupled with a relatively high slippage rate means that some MS slippage is expected in a cancer. Consequently, some authors have questioned whether MSI-L cancers do in fact form a pathological group that is distinct from cancers that show no MS slippage (LAIHO *et al.*, 2002), and have suggested that the MS slippage observed in MSI-L is due to infrequent random events during tumorigenesis, rather than representing qualitatively different genetic pathways (HALFORD *et al.*, 2002).

A previous statistical analysis of MSI frequencies have not found evidence for a distinct MSI-L classification, and suggested that the data is best explained by just MSS and MSI-H classes. Gonzalez-Garcia and colleagues compared how well data from $n = 415$ patients genotyped at $N = 6$ loci was explained by a model with either just MSS cancers, MSI-H and MSS cancers, or MSI-L, MSI-H and MSS cancers (GONZÁLEZ-GARCÍA *et al.*, 2000). In each model, the number of mutated loci expected in a cancer was modelled as binomial distribution, dependent on the mutation rate r. r depended on whether the tumour was MSS and MSI-L and MSI-H. The best fit of the model was given by the two population model where there were just MSI-H and MSS cancers. HALFORD *et al.* (2002) noted that their cohort of cancers showed an excess of MSS cancers (no 'unstable' loci) and a slight excess of cancers with between 10-25% of instability. In their study that genotyped $n = 90$ cancers at $N = 377$ MS loci, LAIHO *et al.* (2002) noted that the proportion of markers unstable in a cancer appeared to increase with patient age. Thus, they conclude that the slight excess of MS slippage

observed in some cancers could just reflect an increase in the number of cells divisions during tumorigenesis.

In this chapter, a mathematical model of MS slippage during carcinogenesis was developed. The aim of the model was to predic the degree of MS slippage that was expected in a cancer from a low rate of normal somatic slippage. The model was then used to assess the degree of MSI observed in a cohort of non-MSI-H colorectal cancers, and decide whether the low-level of microsatellite slippage observed in some of these cancers could be reasonably attributed to normal somatic slippage, or whether an increased MS slippage rate was required. In contrast to the previous techniques described above, the approach taken here describes both the degree of instability present at each locus, as well as the considering the number of unstable loci. This permits a more precise characterisation of the degree of apparent genetic instability present in a cancer.

6.4 Methods

6.4.1 Laboratory methods

MS slippage was examined in 47 paraffin embedded sporadic colorectal cancers. Permission to perform molecular analysis of these cancers was granted by the NHS Oxfordshire Research Ethics Committee B. Each cancer was genotyped using a panel of $N = 9$ MS loci as listed in Table 6.1. The panel consisted of seven dinucleotide repeats and two tetranucleotide repeats. All markers were located either in intronic regions of genes, or outside of genes. Hence, slippages at these loci presumably had little, if any, selective effect on a cell. All loci had been previously identified as prone to slippage in non-MSI-H cancers (HALFORD *et al.*, 2002, 2003).

Macro-dissection

Genotyping experiments were performed by Sarah Halford (HALFORD, 2005). Briefly, each cancer was macro-dissected into four distinct regions using a sterile needle. Regions were selected to be easily macro-dissected, so as to minimise contamination between areas. DNA was extracted according to the macro-dissection protocol

Table 6.1: Details of genotyped microsatellite loci.

Locus name	Location	Repeat type
D10S197	Intron of GAD2	Dinucleotide
D13S175	Not in gene	Dinucleotide
D15S144	Not in gene	Tetranucleotide
D15S659	Not in gene	Tetranucleotide
D17S250	Not in gene	Dinucleotide
D18S58	Not in gene	Dinucleotide
D2S123	Not in gene	Dinucleotide
D5S346	Intron of REEP5	Dinucleotide
D8S87	Not in gene	Dinucleotide

described in Section 4.4.1.

Genotyping

PCR was performed using standard reaction conditions. PCR products were analysed with the ABI377 prism sequencer (Perkin-Elmer), and results were interpreted using Genotyper software (Perkin-Elmer). If MS slippage was identified at a particular locus, the genotyping was repeated to confirm that the slippage was not the result of slippage during PCR amplification. The results of the genotyping are shown in full in the Appendix, Section 8.4.

MSI-H status

MSI-H status of each cancer was determined by presence or absence of slippage at the BAT-26 locus (STONE *et al.*, 2000).

6.4.2 Descriptive statistic

A collection of allelic lengths could be present in each PCR product. To describe the alleles, the *average minimum mean-squared distance from the germline* (denoted by

A), was calculated. The molecular methods used to genotype the cancers are unable to determine which of the germline alleles each allele in the sample was derived from. Hence the statistic A was designed to assume *maximum parsimony*, such that, an allele of length l was derived from the germline allele that was the most similar in size.

Analysis of the PCR product in the Genotyper software indicated the relative concentration of each allele in the PCR product. However, these measurements tend to be unreliable as an indicator of the true proportions of each allele in the sampled cancer cells, due to possible contaminating normal tissue and slippage that occurs during PCR (SIEBEN *et al.*, 2000). Hence the statistic A considered only whether an allele of a particular size was present or absent in the PCR product.

For each region of each cancer, the lengths of the alleles present in the sample, at the locus i, were denoted by l_{i1}, l_{i2}, ..., l_{in}. The germline lengths were at locus i were denoted by g_{i1} and g_{i2} respectively. Then the average minimum mean-squared distance from the germline at locus i (a_i) was defined as:

$$a_i = \left(\frac{1}{n} \sum_{j=1}^{n} \min \left((l_{ij} - g_{i1})^2, (l_{ij} - g_{i2})^2 \right) \right)^{\frac{1}{2}} \tag{6.1}$$

The data from all N loci were summarised as:

$$A = \frac{1}{N} \sum_{j=i}^{N} a_i \tag{6.2}$$

6.4.3 Microsatellite slippage model

The mathematical model of MS slippage during cancer growth was based on that of TSAO *et al.* (2000) and is summarised in Figure 6.2. The model described slippage at a MS locus before and during cancer growth. It was assumed that a cancer grew from a single rogue cell. The lineage of this cell was simulated from birth until a cancer of approximately 10^9 cells, which corresponds to a cancer of about 1cm^3 in volume, was formed. It was assumed that cancers grew exponentially after T_G divisions, where T_G was chosen according to the patients age at biopsy so that cancers had been growing for about ten years before being removed from the body. Cells were assumed to divide at a rate of $C = 182$ divisions per year, which corresponded to a cell division approximately every two days. The rate of cell division was assumed to remain constant during cancer growth.

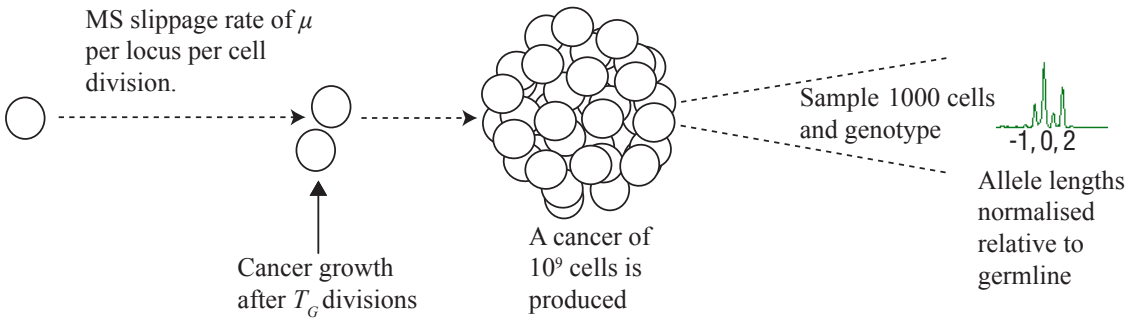


Figure 6.2: Summary of mathematical model of microsatellite slippage during tumorigenesis. The lineage of a cell is simulated from birth until a cancer of approximately 10^9 cells is produced. Cancer growth begins after T_G divisions, where T_G was chosen so that a cancer had been growing for 10 years at biopsy. Microsatellite slippage of ± 1 motif repeat occurs at a constant rate μ at each cell division; cells divide 182 times per year. After the cancer growth had been simulated, a sample of $S = 1000$ cells was taken from the cancer, and the MS alleles present in the sample analysed. Allele lengths were normalised relative to the germline. An allele was considered detectable in the sample if it constituted more than $X\%$ of the sample, where X was the PCR detection threshold.

Cancer growth rates

Rather than growing exponentially, cancers may expand much more slowly (linearly), especially if there is a high rate of cell death within the cancer. This alternative slower growth would mean that there were more cell divisions separating the cells in the cancer from their most recent common ancestor, and hence more variation would be expected at each MS locus. Hence the construction of the model was designed to produce a conservative estimate of the degree of MSI expected in a cancer.

Microsatellite slippage

Pre-cancer cells were assumed to be diploid, so contained two allelic copies of a MS locus. The initial normalised germline lengths of the locus were denoted as g_1 and g_2 respectively where, without loss of generality, $g_1 = 0$ and $g_2 \geq g_1$.

A basal model of stepwise microsatellite slippage was assumed. Each time a cell divided, slippage of ± 1 motif repeats could occur at each allele of the MS locus

with probability μ. The rate of slippage remained constant throughout the simulation, and was assumed to be the same for all MS loci. Somatic MS slippage is relatively poorly characterised so this simple stepwise model was chosen as an adequate first approximation to the presumably complex patterns of MS slippage in colorectal cancers (LAI and SUN, 2003). This simple model was also conservative with respect to the degree of slippage expected under the null hypothesis, since slippages of a more than one motif repeat were not permitted.

Published estimates of MS slippage rates in cancer cells were very broad (HANFORD *et al.*, 1998; BHATTACHARYYA *et al.*, 1994), and rates appeared to depend upon the number of motif repeats present at a particular locus (LAI and SUN, 2003). Hence, the model was examined with two relatively low slippage rates: $\mu = 5 \times 10^{-5}$ and $\mu = 10^{-5}$ per cell per division. These relatively low rates compared favourably with the rates used in other published models (TSAO *et al.*, 2000). The rates were chosen not be a definitive estimate of the MS slippage rate at each locus, but rather to allow the effect of low background rate of somatic slippage to be examined.

Cell ploidy

To examine the effect of cell ploidy, it was assumed that cancer cells were either diploid or tetraploid. In the tetraploid case, it was assumed that cancer cells became tetraploid due to abnormal chromosome duplication at the initiation of cancer growth, that resulted in two copies of each of the two original alleles. Following this initial aberrant event, it was assumed that no further abnormal chromosome duplication occurred. Slippage at each of the resulting four alleles then occurred according to the stepwise mutation model described above.

PCR Detection threshold

Once cancer growth had been simulated, a number of cells ($S = 1000$) were randomly sampled from the cancer. Due to the inherent variability of the PCR process, an allele present in a sample would only be reliably identified if it constituted a significant proportion of the total alleles in the sample (SIEBEN *et al.*, 2000). To account for this, in the model the *PCR detection threshold* was set at $X\%$, so that an allele of length k

was only considered present in the sample if at least $X\%$ of the total number of alleles in the sample had length k. The model was simulated for two plausible dectection thresholds of $X = 10\%$ and $X = 20\%$.

6.4.4 Model construction and simulation

A coalescent type construction was used for the model, similar to that of TSAO *et al.* (2000). Using this framework, only the ancestry of the $S = 1000$ cells that were sampled from the cancer needed to be simulated, rather than the ancestry of the entire 10^9 cells in the cancer.

The simulation had three parts:

1. Offspring counts. The number of cells having 0, 1 or 2 offspring in each generation were computed. In generation i, the number of cells having 0, 1 or 2 offspring were labelled n_{i0}, n_{i1} and n_{i2} respectively, and were computed for each generation of cells from the zygote until the cancer of 10^9 cells was formed after age $\times\, C$ divisions (see Table 6.2). Cancer growth began after T_G, where T_G was chosen such that a cancer grew for a period of ten years, thus $T_G = (\text{age} - 10)C$. Prior to T_G divisions, each cell had a single offspring cell that formed part of the lineage of the tumour, hence $n_{i0} = n_{i2} = 0$ and $n_{i1} = 1$ for $i = 1, \ldots, T_G$. After T_G divisions, the cancer was assumed to grow exponentially such that a cancer of 10^9 cells was formed after ten years. Note that the number of cells in the ith generation is then $n_{i1} + 2n_{i2}$.

2. Construction of ancestral tree of sampled cells. The offspring counts computed in part 1 were used to construct the ancestral tree that defines the relations between cells in the sample. A sample of $S = 1000$ cells was "sampled" at random from the cancer. Suppose that the sample was taken from the offspring of generation i. The offspring counts from generation i defined the probability of each of the cells in the sample sharing a parent cell in generation $i - 1$. Thus, for $i = \text{age} \times 182, \ldots, 1$, the ancestral relations between each cell in the sample could be constructed.

3. Simulation of microsatellite slippage. The initial cell, at generation $i = 0$

was initialised with the germline microsatellite lengths g_{i0} and g_{i1}. Starting at generation $i = 1$ and moving the down the ancestral tree constructed in step 2, the microsatellite lengths of each daughter cell in the tree were simulated according to the microsatellite slippage model described above (Section 6.4.3).

Table 6.2: Example of computation of number of offspring cells in each generation.

	Numbers of cells			
Generation	**0 offspring**	**1 offspring**	**2 offspring**	**Total cells in next generation**
0	0	1	0	1
1	0	1	0	1
2	0	1	0	1
3	0	1	0	1
4	0	1	0	1
5	0	1	0	1
6	0	1	0	1
7	0	1	0	1
8	0	1	0	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
T_G	0	0	1	2
T_G + 1	0	1	1	3
T_G + 2	0	1	2	5
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Age/182	0	100000000	450000000	1000000000

Simulation of model

The model was written in C. Simulations were performed on the UCL Central Computing Services C^3 Linux cluster. The simulation results were compared to the laboratory data using R (R DEVELOPMENT CORE TEAM, 2007).

6.4.5 Comparison of model to data

To compare the expected versus observed MS slippage the statistic A was used. For each patient, the model was simulated for each of the N MS loci. These N simulations were described as a single *instance* of the model. An approximation to the underlying distribution of A (labelled $\widehat{A}$) that was expected under the model was produced by repeated simulation of the model, 10^5 instances for each cancer, and recording each value of A. The distribution $\widehat{A}$ depended upon the germline allele lengths of each MS locus and the age of the patient at biopsy, and so had to be simulated for each cancer in turn.

To calculate a p-value, the probability that MS slippage at least as extreme as the slippage observed in each region of the cancer would occur under the 'null hypothesis' model, the proportion of simulation repeats that produced a value of A at least as extreme as observed in the cancer region was calculated. This p-value was defined as:

$$p = P(A_{sim} \geq A_{obs}) \tag{6.3}$$

where A_{sim} was the simulated value of A and the A_{obs} was the value of A observed in the cancer. For each sample, the null hypothesis was rejected if $p < 0.05$ and the cancer was labelled putative MSI-L.

6.5 Results

The microsatellite slippage observed in a cohort colorectal cancers was assessed using a model of normal somatic microsatellite slippage during tumorigenesis. Comparison of the degree of slippage predicted by the model, to the degree of slippage observed in each cancer, allowed the testing of the 'null hypothesis' that the MS slippage observed in the cancer was due only to a low rate of background slippage.

6.5.1 Model behaviour

Effect of detection threshold

Increasing the PCR detection threshold from 10% to 20% did not significantly alter the distribution of A values (see Figure 6.3).

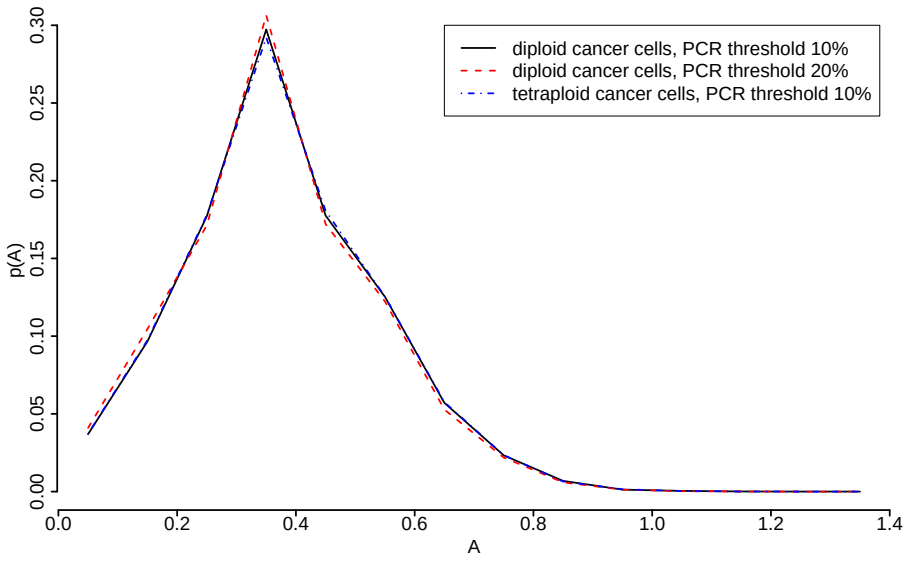


Figure 6.3: Effect of altering cell ploidy or PCR detection threshold. Simulations were computed with $N = 9$ loci, germline lengths $g_0 = 0$, $g_1 = 1$, $\mu = 5 \times 10^{-5}$, with a PCR detection threshold $X = 10\%$ and diploid cancer cells unless otherwise stated.

Effect of cancer cell ploidy

Assuming cancer cells were tetraploid or diploid, due to an erroneous cell division at the initiation of cancer growth did not significantly alter the p-values (see Figure 6.3).

Number of loci successfully genotyped

The variance of A (σ_A) increased as the number of successfully genotyped loci (N) decreased (see Figure 6.4). As N increased, the variance of A monotonically decreases.

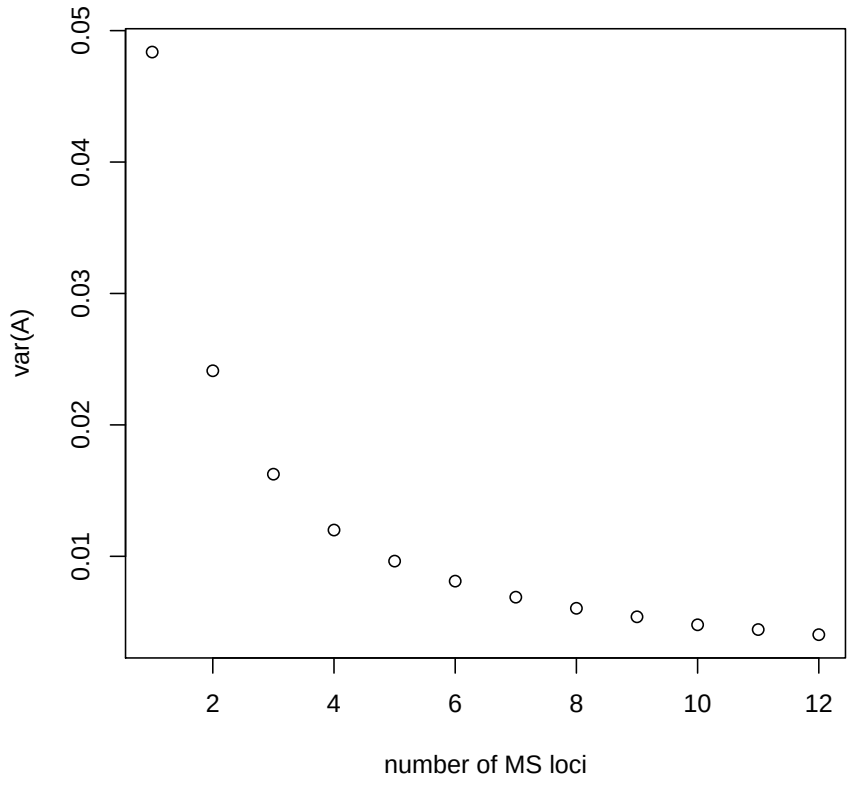


Figure 6.4: Variance of A as a function of N, the number of successfully genotyped loci. The variance of A decreases as the number of successfully genotyped loci increases. Simulations were computed with germline lengths $g_0 = 0$, $g_1 = 1$, a PCR detection threshold $X = 10\%$, $\mu = 10^{-5}$ and diploid cancer cells.

Slippage rate

Increasing the slippage rate (μ) increased the mean value of A (see Figure 6.5.1), so that more MS slippage was expected in a cancer due to normal somatic slippage.

6.5.2 Validation of model on MSI-H cancers

The model was compared to the slippage observed in the five MSI-H cancers (See Table 6.5, cancers H1-H5). This comparison served as a 'test-case', to see if the model of normal somatic slippage would be unable to explain the slippage observed in cancers with an elevated slippage rate.

Five MSI-H cancers in the cohort of 47 cancers were identified by the presence of

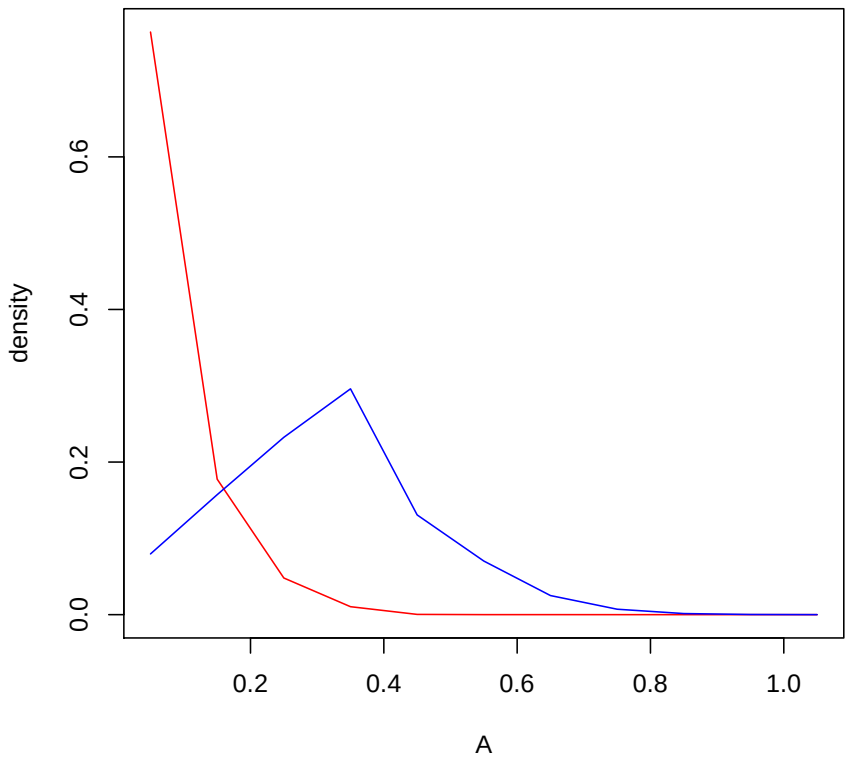


Figure 6.5: Distribution of A for $\mu = 10^{-5}$ (red line) and $\mu = 5 \times 10^{-5}$ (blue line). Simulations results are for $N = 9$ loci, each with germline lengths $g_0 = 0$, $g_1 = 1$, PCR detection threshold $X = 10\%$ and diploid cancer cells.

slippage at the BAT-26 locus (STONE *et al.*, 2000). All five MSI-H cancers had significant MS slippage at most of the loci examined. The slippage rate was set at $\mu = 10^{-5}$ per locus per cell division, and it was assumed that cancer cells were diploid and that the PCR detection threshold was 10%. Then, in all regions of the five cancers the degree of MSI exhibited was very unlikely to occur under the null hypothesis. P-values are shown in Table 6.5, and the largest p-value observed was $p < 10^{-5}$. Similar results were obtained if the slippage rate was increased to $\mu = 5 \times 10^{-5}$ (largest p-value: $p = 0.026$). As noted above, cancer cell ploidy and the PCR detection threshold did not significantly alter the p-values. Therefore it was confirmed that the MS slippage in the five MSI-H cancers was very likely to have resulted from an elevated slippage rate.

Table 6.3: Values of A and calculated p-values for each cancer (Cancers 1-17). P-values were computed with $\mu = 10^{-5}$, diploid cancer cells and a PCR detection threshold of $X = 10\%$. Significant p-values and corresponding values of A are highlighted in **bold**.

ID	Age at biopsy	MSI-H	N	Sex	Site	Dukes	A1	A2	A3	A4	P1	P2	P3	P4
1	23	no	8	F	rectosigmoid	B	0.072	0.072	0.072	0.072	0.338	0.338	0.338	0.338
2	50	no	9	F	rectum	C	0	0	0	0	1.000	1.000	1.000	1.000
3	53	no	9	F	sigmoid	B	0	0	0	0	1.000	1.000	1.000	1.000
4	54	no	7	M	rectum	-	**0.558**	**0.325**	**0.248**	0.143	**0.000**	**0.011**	**0.046**	0.384
5	55	no	9	F	rectum	B	0	0	0	0	1.000	1.000	1.000	1.000
6	56	no	9	M	rectum	-	0	0	0.192	0	1.000	1.000	0.171	1.000
7	57	no	9	M	rectum	B	0	0	0	0	1.000	1.000	1.000	1.000
8	58	no	9	M	rectum	B	0	0	0.2	0	1.000	1.000	0.160	1.000
9	58	no	9	F	ascending	C	0	0	0	**0.436**	1.000	1.000	1.000	**0.002**
10	58	no	9	F	rectum	A	**0.356**	0	0.128	0	**0.011**	1.000	0.447	1.000
11	59	no	8	F	rectum	B	0	**0.451**	0	0	1.000	**0.002**	1.000	1.000
12	60	no	9	F	sigmoid	B	0.064	0	0	0	0.810	1.000	1.000	1.000
13	63	no	9	F	Hepatic Flexure	A	0	0	0	0	1.000	1.000	1.000	1.000
14	64	no	9	F	rectum	-	0	0	0	0	1.000	1.000	1.000	1.000
15	64	no	9	M	rectum	B	0.064	0.064	0.064	0	0.817	0.817	0.817	1.000
16	64	no	9	M	sigmoid	C	0	0	0	0	1.000	1.000	1.000	1.000
17	64	no	8	M	sigmoid	B	0	0	0	0	1.000	1.000	1.000	1.000

Table 6.4: Values of A and calculated p-values for each cancer (Cancers 18-33). P-values were computed with $\mu = 10^{-5}$, diploid cancer cells and a PCR detection threshold of $X = 10\%$. Significant p-values and corresponding values of A are highlighted in **bold**.

ID	Age at biopsy	MSI-H	N	Sex	Site	Dukes	A1	A2	A3	A4	P1	P2	P3	P4
18	64	no	8	M	rectum	A	**0.357**	0	0.072	**0.422**	**0.008**	1.000	0.727	**0.005**
19	65	no	9	M	rectum	-	0.128	0.192	0	0.064	0.458	0.185	1.000	0.801
20	65	no	6	F	sigmoid	A	**0.3**	**0.375**	0.186	**0.31**	**0.047**	**0.011**	0.210	**0.047**
21	65	no	9	M	rectum	C	0	0.124	0.124	0	1.000	0.520	0.520	1.000
22	65	no	9	M	sigmoid	C	0.15	0.086	0.15	0.122	0.485	0.503	0.485	0.487
23	66	no	9	F	sigmoid	B	0	0	0	0	1.000	1.000	1.000	1.000
24	67	no	9	M	ascending	C	0	0	0	0	1.000	1.000	1.000	1.000
25	67	no	9	M	descending	C	0	0	0	0	1.000	1.000	1.000	1.000
26	70	no	9	M	sigmoid	-	0	0	0	0	1.000	1.000	1.000	1.000
27	70	no	9	F	sigmoid	B	0.064	0.064	0.064	0.064	0.845	0.845	0.845	0.845
28	71	no	6	M	rectum	C	0.224	0.115	0.231	0	0.237	0.634	0.237	1.000
29	71	no	6	M	sigmoid	A	**0.542**	**0.77**	**0.493**	**0.577**	**0.003**	**0.000**	**0.005**	**0.002**
30	72	no	9	M	descending	C	0.124	0.124	0.124	0.124	0.548	0.548	0.548	0.548
31	73	no	9	F	sigmoid	C	0	0	0	0	1.000	1.000	1.000	1.000
32	75	no	9	M	rectum	-	0	0	0	0	1.000	1.000	1.000	1.000
33	75	no	9	M	rectum	B	0	0	0	0	1.000	1.000	1.000	1.000

Table 6.5: Values of A and calculated p-values for each cancer (Cancers 34-H5). P-values were computed with $\mu = 10^{-5}$, diploid cancer cells and a PCR detection threshold of $X = 10\%$. Significant p-values and corresponding values of A are highlighted in **bold**.

ID	Age at biopsy	MSI-H	N	Sex	Site	Dukes	A1	A2	A3	A4	P1	P2	P3	P4
34	78	no	9	F	rectum	B	0	0	0.144	0	1.000	1.000	0.523	1.000
35	78	no	9	F	rectum	C	0	0.064	**0.317**	0	1.000	0.860	**0.038**	1.000
36	79	no	5	M	ascending	C	**0.488**	0.365	0.365	0.341	**0.012**	0.071	0.071	0.091
37	80	no	9	M	ascending	B	0	0	0	0	1.000	1.000	1.000	1.000
38	80	no	9	F	sigmoid	C	0	**0.411**	0.314	0	1.000	**0.015**	0.120	1.000
39	81	no	9	M	rectum	A	0	0	0	0	1.000	1.000	1.000	1.000
40	83	no	9	F	rectum	C	0.124	0.143	0.124	0.218	0.662	0.661	0.662	0.370
41	84	no	9	F	rectum	B	0.229	0	0.229	0	0.309	1.000	0.309	1.000
42	86	no	9	M	rectum	B	0	0	0	0	1.000	1.000	1.000	1.000
H1	63	yes	7	F	ascending	B	**2.016**	**2.014**	**1.966**	**1.829**	**0**	**0**	**0**	**0**
H2	67	yes	7	F	sigmoid	B	**1.489**	**1.489**	**1.489**	**1.553**	**0**	**0**	**0**	**0**
H3	69	yes	9	F	sigmoid	C	**0.816**	**0.816**	**0.816**	**1.2**	**0**	**0**	**0**	**0**
H4	75	yes	8	M	sigmoid	C	**0.91**	**0.913**	**0.807**	**0.788**	**0**	**0**	**0**	**0**
H5	103	yes	9	F	ascending	C	**2.141**	**1.933**	**1.696**	**1.913**	**0**	**0**	**0**	**0**

6.5.3 MS slippage in non-MSI-H cancers

Next the model was compared to the slippage observed in the 42 remaining non-MSI-H cancers. Initially a slippage rate of $\mu = 10^{-5}$ per locus per cell division was assumed. Cancer cells were assumed to be diploid and the PCR detection threshold was set at 10%. Using these assumptions, 10 out of the 42 non-MSI-H cancers (24%) contained at least one region where the degree of MSI was too severe to be explained by the null hypothesis ($p < 0.05$ in these cancers) suggesting that these 10 cancers had acquired an increased slippage rate during tumorigenesis and were indeed MSI-L (see Tables 6.3-6.5). Borderline levels of MSI ($p < 0.1$) were observed in regions of two cancers (cancers 36 and 38), but both these cancers contained another region where the amount of slippage was indicative of MSI-L.

The degree of MSI in these ten putative MSI-L cancers was then examined (see Tables 6.3-6.3). 4 of these 10 cancers (cancers: 9-11 and 38) had a putative MSI-L region that had only one unstable locus, which always had multiple extra alleles that showed a maximum deviation from the germline of at least five motif repeats. 5 cancers had MSI-L regions with two unstable loci (cancers: 4, 18, 20, 29, 36) and only one cancer had a MSI-L region with three unstable loci (cancer 35); these cancers had multiple extra sizes of alleles at at least two loci, or had at least two unstable loci, one of which that had an allele that differed by at least three motif repeats from the germline. In comparison, non-MSI-L cancers always had less than three regions showing slippage, which was less severe than in putative MSI-L cancers. Typically a value of $A > 0.25$ was sufficient to produce a p-value less than 0.05. Cancer 36 had values of $A = 0.37$, 0.37 and 0.34 in regions 2-4, respectively, but these values did not produce a significant p-value. This is because only five loci were successfully genotyped in cancer 36, and the variance of A is greater when fewer loci are examined (see Figure 6.4).

6.5.4 Most microsatellite slippage in non-MSI-H cancers can be explained by background slippage

Next, the data were compared to the model with a slightly higher slippage rate of $\mu = 5 \times 10^{-5}$ per locus per cell per division. It was again assumed that cancer cells were diploid and that PCR detection threshold was 10%. With this slightly higher slippage

rate, the slippage observed in all the non-MSI-H cancers could be adequately explained by the null hypothesis. The smallest p-value observed in the non-MSI-H cancers was $p = 0.094$.

6.5.5 Associations with clinico-pathological variables

Dukes stage (Fisher's Exact test, $p = 0.7115$), location in colorectum (Fisher's Exact test, $p = 1$), patient sex (Fisher's Exact test, $p = 0.4682$) and age at biopsy (Mann-Whitney test $p = 0.8015$) was compared between the putative MSS and MSI-L cancers, identified with $\mu = 10^{-5}$. No significant associations were found.

6.5.6 Timing of MSI-L

Of the 10 cancers identified as putative MSI-L, only one (cancer 29) appeared to have MSI-L in all four regions. Cancer 29 had two unstable loci that showed significant slippage in all four regions of the cancer and had a maximum change of at least 3 repeats from the germline. Of the remaining 9 putative MSI-L cancers, 5 cancers had MSI-L in only one region, 1 cancer had MSI-L in two regions, and a further 2 cancers had MSI-L in three regions. Therefore, if these 9 cancers were truly MSI-L, then the increase in slippage rate was likely to have occurred late in the development of the cancer. If MSI-L had occurred early, then all the regions of a cancer would show similarly high degrees of MSI.

6.5.7 Number of unstable loci as a predictor of MSI-L status

It was considered how reliable the number of unstable loci was as a predictor of MSI-L status. The model with $\mu = 10^{-5}$ was assumed to correctly identify regions of a cancer that were MSI-L. Then the sensitivity and specificity of identifying these regions by their having at least 1, 2 or 3 loci that showed slippage was calculated (see Table 6.6). Classifying a cancer as MSI-L by the presence of slippage at a single locus had poor specificity, whereas requiring slippage at multiple loci for MSI-L classification had poor sensitivity.

Table 6.6: Sensitivity and specificity of using the number of unstable loci for MSI-L classification. The model with slippage rate $\mu = 10^{-5}$ was assumed to identify reliably regions of each cancer that were MSI-L. The sensitivity and specificity of counting the number of unstable loci in each region was then computed.

Minimum number of unstable loci required for MSI-L classification	**Sensitivity**	**Specificity**
1	1 (18/18)	0.69 (102/150)
2	0.72 (13/18)	0.97 (145/150)
3	0.06 (1/18)	1 (150/150)

6.6 Discussion

Using a mathematical model of microsatellite slippage during tumour growth, the degree of microsatellite slippage expected in a cancer without an elevated slippage rate was predicted. Assuming that the normal somatic slippage rate is relatively low (slippage rate of $\mu = 10^{-5}$ per locus per cell per division) then most slippage observed in a group of non-MSI-H cancers could be reasonably explained by normal somatic slippage. A minority of the non-MSI-H cancers that were genotyped (24%) contained a region that had slippage that was more severe than predicted by the model, and so were labelled as MSI-L. If the proportion of the regions that showed 'severe' slippage was considered, then the frequency of putative MSI-L regions was only 18/168 (11%). Hence, if the model did correctly identify MSI-L cancers, then the model suggested that MSI-L cancers were relatively rare.

Only a single cancer was observed to have a degree of MSI that was indicative of MSI-L in all four regions that were genotyped. This suggests that the acquisition of an MSI-L phenotype is not an early event in tumorigenesis, since it is likely that an increase in the slippage rate early in tumorigenesis would cause detectable slippage in most of the cells in the cancer.

If the normal MS slippage rate was assumed to be slightly higher, $\mu = 5 \times 10^{-5}$

per locus per cell per division, all of the slippage observed in the non-MSI-H cancers could be adequately explained by background slippage. With this elevated mutation rate, slippage in MSI-H cancers could still not be explained by the model. Although precise estimates of somatic slippage rates were unavailable, estimates of MS slippage rates between human generations, and slippage rates in MMR-proficient cancer cells, ranged from 10^{-8}-10^{-3} per locus per division (LAI and SUN, 2003; HANFORD *et al.*, 1998; KRUGLYAK *et al.*, 1998), and were typically toward the upper end of this scale (KRUGLYAK *et al.*, 1998; LAI and SUN, 2003), so an average background rate of $\mu = 5 \times 10^{-5}$ is reasonable. Moreover the $N = 9$ MS loci examined in this study were selected as they were sensitive for detection of putative MSI-L cancers (HALFORD *et al.*, 2002, 2003), so they were likely to have a relatively high slippage rate.

Although it was not possible to be sure of the precise background somatic MS slippage rate at each locus, the model illustrated that, given a low somatic slippage rate, it was reasonable to expect a non-negligible degree of slippage in a cancer. It should be noted that it is unlikely that the model could accurately predict the true degree of MSI that would occur in any particular cancer. This does not detract from the analysis presented here. The model was designed only to provide a prediction of the basal level of microsatellite slippage that could be expected in any cancer, so that the prediction could be used to critically assess the degree of MSI observed in a cohort of non-MSI-H cancers. The model clearly illustrated that occasional 'MSI' is expected from normal somatic slippage alone, and so low-level microsatellite slippage is not necessarily the result of some underlying change that causes an increase in the slippage rate.

Furthermore, the construction of the model was designed to minimise the degree of MS slipapge expected by chance, and so the predictions of the model are conservative with respect to the existence of MSI-L. For example, in the model, cancers were assumed to grow for 10 years, whereas recently published data suggests that cancers may actually grow for more than 17 years (JONES *et al.*, 2008b). Also, it was supposed that cancers grew at a slow exponential rate, whereas growth may be much slower, perhaps linear, so that there would be more rounds of cell division in which slippages can accumulate. Further, it was assumed that cancer cells divide about once every two days. There are many estimates of cancer cell division rates, but notably, similar work

has assumed a rate of one cell division per day (TSAO *et al.*, 2000; YATABE *et al.*, 2001; BEERENWINKEL *et al.*, 2007). This higher rate of division would mean that more slippage would be expected in a cancer due to normal somatic slippage, since more rounds of cell division and death would occur in the ten years that a tumour had been growing. Therefore, it is concluded that an elevated slippage rate is not required to explain most MS slippage observed in non-MMR-deficient cancers.

Analysis of the number of 'unstable' loci as an indicator of MSI-L status was shown to be unreliable. The model suggests that cancers which have at least one locus that shows significant slippage (a change of at least 4 motif repeats from the germline and multiple extra alleles), or multiple 'unstable' loci, at least one of which is highly unstable, are the most likely to have acquired a mutator phenotype. An approximately equivalent indicator is a value of $A > 0.25$ when sufficient loci ($N \geq 9$) are examined.

Does MSI-L exists even in a minority of non-MSI-H cancers? During PCR, MS loci can slip (SIEBEN *et al.*, 2000), so MSI will be occasionally observed in cancers that are actually MSS. MSI-L could be the result of external mutagenic factors, so that occasional MS slippages result from unpredictable events in that occur during the development the cancer, rather than an intrinsic genetic or epigenetic defect. A low slippage rate, coupled with the many rounds of division that occur during the growth of a cancer, means that some slippage is expected in a cancer even without an elevated slippage rate. Clear evidence for a distinct MSI-L group of cancers would be provided by the discovery of a strong association of MSI-L with a clinical or pathological variable. Previous work has looked for such an association, by examining the co-occurrence of putative MSI-L an variables such as K-ras mutations, loss of heterozygosity events at APC, or methylation of O6-methylguanine methyltransferase, as noted in the Introduction to this Chapter. The findings have been inconsistent, with most studies finding no significant association between MSI-L and any clinico-pathological or molecular variable. Potentially, the reason for the inconsistency could be the difficulty in identifying MSI-L cancers. The associations could be re-examined using the definition of MSI-L provided in this chapter.

6.7 Conclusions

There is little evidence for a distinct group of MSI-L cancers. A low background somatic slippage rate can explain most apparent low-level microsatellite instability observed in non-MSI-H colorectal cancers. A few non-MSI-H cancers exhibit a degree of slippage more severe than expected from a low rate of somatic slippage, although the degree of MSI observed in these cancers can be explained by a slightly higher somatic slippage rate. If MSI-L did occur in some cancers, it was not an early event in their development, since the degree of MSI in a cancer varied between spatially distinct regions. Counting the number of unstable loci in a cancer is a poor predictor of MSI-L status.

6.8 Future Work

6.8.1 Association with clinicopathological variables

Previously published work has suggested an association between MSI-L and various genetic, clinical or pathological features of a cancer, such as K-ras mutations (JASS *et al.*, 1999), MGMT hypermethylation (WHITEHALL *et al.*, 2001), cancer staging (WRIGHT *et al.*, 2005) and the level of immune response (KETS *et al.*, 2006). Evidence for the associations is contested though, and as many other studies found no such associations (for example, see review by TOMLINSON *et al.* (2002)). The reason for the inconsistency could be the difficulty in reliably identifying MSI-L cancers. Using the definition of MSI-L suggested in this chapter, evidence for these associations could be re-examined. If associations are relatively weak, a large sample size of cancer will be required to confidently detect an association.

6.8.2 Molecular analysis

Recent developments in sequencing technology permit whole genome sequencing of small tissue samples (ILLUMINA, 2006). The frequency distribution of the number of motif repeats per MS locus (constructed from published copies of the human genome) have been used to estimate the MS slippage rate between human generations (see for example LAI and SUN (2003); KRUGLYAK *et al.* (1998)). The frequency distribution

is assumed to represent an 'equilibrium distribution' such that the frequency of a locus consisting of X motif repeats is maintained by balance between contraction of longer loci and expansion of shorter loci. The slippage rate is estimated by fitting a stochastic model of MS expansion and contraction to this equilibrium distribution. Assuming that the distribution of microsatellite lengths in cancers is approximately in equilibrium, this method could be applied to study somatic slippage rates during cancer growth. First, a mathematical analysis would be required to show that the distribution of MS lengths in a cancer was in quasi-equilibrium. Whole genome PCR, with deep re-sequencing, could be used to measure the frequency distribution of the number of motif repeats per MS locus in MSI-H, MSS and putative MSI-L cancers, and similar mathematical techniques could be used to infer the microsatellite slippage rate in each cancer type. This analysis would provide a powerful assay to detect MSI-L cancers, and would also result in an estimate of the somatic MS slippage rate. These data could also be used to identify a panel of non-polymorphic microsatellites, which show slippage below the level expected from somatic slippage. MSI-L cancers may then be sensitivity identified if they showed slippage at the MS in this panel.

6.8.3 Model complexity

The model of MS slippage during tumorigienesis presented in this chapter is intentionally simple. The intention of the work presented in this chapter was to use a simple model of normal somatic MS slippage during tumorigenesis to predict the basal level of microsatellite slippage that could be reasonably expected in a cancer without an elevated slippage rate. However, it is unlikely that this parsimonious description is an accurate representation of the complex factors that influence MS slippage during tumorigenesis. A more accurate model should assess the following assumptions and introduce more complex models as appropriate: per locus slippage rates, slippage rates that depend on the number of motif repeats at the locus, different cancer growth rates and heterogeneous cell division rates.

6.8.4 CIMP-low

Some authors have suggested that a sub-group of cancers can be identified by their having a low level *CpG Island Methylator Phenotype* (CIMP-Low) (OGINO *et al.*,

2006; JASS, 2007; KAWASAKI *et al.*, 2008). It is suggested that in these cancers, CpG island methylation occurs at a slightly higher rate than the background somatic methylation rate. Methylation of a tumour suppressor gene, or its promoter, which suppresses protein transcription can lead to neoplastic development. Hence, the acquisition of CIMP-low predisposes to cancer development in an analogous way to the acquisition of the MSI-L phenotype.

The existence of a CIMP-low phenotype remains controversial, since the number of methylated loci required for CIMP-low classification is chosen arbitrarily. The method adopted in this Chapter could be adapted to assess the evidence for a CIMP-low phenotype. The model developed in this chapter could be adapted to predict the level of methylation expected in a cancer given a low somatic methylation rate. The predictions of the model could then be compared to the level of methylation observed in a cohort of putative CIMP-low cancers.

Chapter 7

Conclusions

The aim of this thesis was to investigate the propagation of mutations during tumorigenesis, particularly in the human colon. First, the expansion of amutant clone in a stem cell niche was examined. The Drosophila testis, a very well characterised niche, was used as a model system to investigate stem cell survival. Next, the propagation of mutations through the colon was examined. The rate of crypt fission, the primary mechanism for expansion of mutant clones in the colon, was estimated. The order and timing of APC and K-ras mutations in the initial development of colorectal adenomas was then investigated. Laser-capture micro-dissection, coupled with optimised sequencing techniques, was used to produce clonal maps to study the expansion of the mutant clones. Finally, mutations in colorectal cancers were investigated. It had been suggested that some cancers could be classified according to their having a low-level mutator phenotype, which caused an elevated rate of microsatellite slippage. This hypothesis was investigated by developing a mathematical model that predicted the level of microsatellite slippage expected in a cancer from normal somatic slippage, and comparing this prediction to the slippage observed in a cohort of colorectal cancers.

The earliest tumorigenic mutations are thought to occur in stem cells, since only stem cells have lineages long enough to be likely to acquire sufficient mutations to initiate tumour growth (PIERCE and SPEERS, 1988; CAIRNS, 2002). To study the initial events in tumorigenesis, in this thesis the clonal expansion of mutant stem cells in a niche was examined. Using the well characterised Drosophila testis stem cell niche as a model

system, a mathematical model of the stem cell compartment was developed and used to identify the stem cell phenotypes that were selectively advantageous. The rate of invasion of stem cells with particular genetic mutations was predicted. This work has generated testable predictions which may stimulate future research into the dynamics of stem cells in the Drosophila testis.

Furthermore, the model illustrated how specific bio-mechanical properties of a stem cell can alter the likelihood that the progeny of the stem cell will remain in a niche. This study of the Drosophila niche should be viewed as an illustration of how an tumorigenic mutation may become fixed in a niche, through subtly altering the phenotype of a cell. After fixation of the mutant cell, the cells in the niche are predisposed to tumorigenesis, even if the original mutation in isolation was insufficient to initiate tumour growth. Studying *pre-tumour progression* in the human colon is prohibitively difficult. Sequencing technology restricts the ability to genotype cells from a single crypt, since the volume of DNA available is small. Recently however, advances in sequencing technology have allowed whole genome sequencing of colorectal neoplasms (JONES *et al.*, 2008b). Adenomas were found to contain between 6 and 23 mutations. This relatively high mutation burden in colorectal adenomas perhaps suggests that some tumorigenic mutations occur prior to APC inactivation, and therefore may influence subsequent tumour progression. Understanding the mechanisms responsible for the accumulation of mutations in small adenomas, either through neutral drift that can lead to the fixation of a mutation, or mutation driven clonal expansion, is likely to become increasingly important as sequencing technology improves.

The next section of this thesis examined the spread of mutations through the colonic epithelium. A mutant stem cell arising in a crypt can displace the wild-type stem cells from the niche through *niche succession* (YATABE *et al.*, 2001), a process analogous to mutant stem cell invasion in the Drosophila testes. Subsequent expansion of a mutant clone is through *crypt fission* (GREAVES *et al.*, 2006). In this thesis, the rate of crypt fission was inferred from data on the spread of COX-deficient cells in the colon. Fission was found to be an infrequent event, occuring in each crypt about once every 23 years. This was the first estimate of the fission rate that was inferred directly from data

from human colons, rather than by comparison with murine crypts. The estimate was significantly longer than the previous availiable estimate. A mathematical model of the partitioning of stem cells between daughter crypts during fission showed that fission increased the likelihood of a mutant lineage becoming fixed in the epithelium.

Crypt fission is the primary mechanism for expansion of colorectal adenomas (PRESTON *et al.*, 2003), and correspondingly the incidence of crypts in fission is greatly increased in adenomas (WONG *et al.*, 2002; BJERKNES *et al.*, 1997). Therefore, the normal low rate of crypt fission, identified in this thesis, may have evolved as an anti-tumourigenic mechanism which limits the spread of mutant clones in the epithelium. Understanding the mechanism responsible for the increase in fission rate in neoplastic development may prove vital in developing treatments to restrict the growth of colorectal adenomas.

Genetic alterations occurring as part of the initial development of sporadic colorectal adenomas were then examined. Polyps were micro-dissected on a crypt-by-crypt basis, and the mutation status of the APC and K-ras genes in each crypt was determined. K-ras mutations were found to occur frequently in small adenomas, and many adenomas had K-ras mutations in every crypt that was genotyped. This suggested that K-ras mutations may occur very soon after the initiation of tumour growth and rapidly reach fixation in the adenoma. Alternatively, this data could be interpreted as evidence of pre-tumour progression, whereby a crypt had acquired a K-ras mutation before APC inactivation and the subsequent initiation of tumour growth. It was also suggested that specific combinations of APC and K-ras mutations tended to be found together, suggesting that tumour cells with these combinations are bestowed with an optimal selective advantage. This work has resulted in a better characterisation of the initial clonal expansions in colorectal cancer development. Laser capture micro-dissection, coupled with optimised sequencing techniques, has been shown to be an efficient tool for investigating the clonal development of colorectal adenomas. The associations between K-ras and APC mutations merit further investigation to understand how combinations of specific mutation types may define different pathways of tumorigenesis.

The final section of this thesis examined the accumulation of mutations in colorectal cancers. It had been previously shown that some cancers, with intact mismatch repair

machinery, exhibit a low level of microsatellite instability (BOLAND *et al.*, 1998), and these cancers have been labelled MSI-L. Much research has been focused on determining whether this apparent low-level of instability is due to an elevated slippage rate, resulting from an underlying genetic change in the cancer (TOMLINSON *et al.*, 2002). In this thesis, it was shown that a normal low level of somatic microsatellite slippage during tumorigenesis will result in a non-negligible degree of slippage in a cancer. This suggested that the microsatellite slippage observed in the MMR-proficient cancers was due to random slippage events during the development of the cancer, rather than an underlying genetic change. Criteria were then suggested which differentiate between degrees of microsatellite instability in a cancer, and so may be useful for the identification of putative MSI-L cancers. These definitions will be useful for studying any association between the putative MSI-L phenotype and clinical or pathological variables. Finding such an association may have prognostic benefit.

There has been a long running debate as to the relative importance of genetic instability versus clonal expansion in driving tumour progression (TOMLINSON *et al.*, 1996; MOOLGAVKAR and LUEBECK, 2003; NOWAK *et al.*, 2002; RAJAGOPALAN *et al.*, 2003). Genetic instability increases the rate at which new genotypes are produced in a cancer. When coupled with selection for the fittest clones, an elevated mutation rate accelerates the process of evolution that occurs in a cancer, and so the time taken for a normal healthy cell to make the transition to an invasive carcinoma is reduced. Clonal expansion can have an equivalent effect. Clonal expansion of a cell lineage, following the initial inactivation of a tumour suppressor gene, produces many cells which are potential 'targets' for further tumorigenic changes. Thus, in a large clone there are many cells which are susceptible to further tumourigenic mutations, so the transition from a normal to malignant clone can be rapid. This thesis has looked at the relative importance of genetic instability and clonal expansion during tumorigenesis. The study of microsatellite instability in putative MSI-L cancers showed that cancers can consist of genetically diverse populations of cells, without requiring an elevated mutation rate to generate diversity. Selection, rather than mutation, may primarily influence the spectrum of mutations observed in colorectal adenomas. Mathematical modelling, as illustrated in this thesis, can reveal cell phenotypes that are selectively

advantageous, and show whether the observed spectrum of mutations is compatible with the proposed model of tumour progression.

Future research directions following from the work presented in this thesis have been suggested at the end of each Chapter. Here, preferred research directions are summarised.

In Chapter 2, the behaviour of sporadic mutant stem cells in the Drosophila testis stem cell niche was predicted. These predictions can be readily tested in a suitable fly model where mutations (such as apc loss) can be induced in the adult fly testis. Development and testing of these animal models would provide unrivalled insight into the earliest events in tumorigenesis. These data should be analysed in tandem with a more sophisticated mathematical than presented in this thesis, describing, in particular, the behaviours associated with an ageing niche.

In Chapter 3, the role of crypt fission in the propagation of mutant cells in the human colon was investigated. Similar analysis, requiring the availability of adequate data, should be applied to study the rate of crypt or gland expansion in other gastrointestinal tissues, such as the small intestine or stomach. Developing a stochastic, rather than deterministic, model of crypt fission would confirm the results presented here. Such models could also be used to test the hypothesis that crypt fission occurs in response to some (random) intra- or extra-cryptal stimulation, rather being an inevitable event in the life-cycle of a crypt.

In Chapter 5, the order and timing of APC and K-ras mutations in sporadic colorectal adenomas was inferred. This work tentatively indicates that APC and K-ras mutations may not occur independently, in that the type of APC mutation may select for the presence and possibly type of K-ras mutation. This hypothesis should be evaluated using mouse models. Mutant mice with different combinations of APC and K-ras mutations can be generated and compared. The analysis could also be extended to consider the role of other genes, such as *BRAF* and *p53*, in the development of these sporadic polyps.

In Chapter 6, the degree of genetic diversity at microsatellite loci expected in a cancer as a result of normal rate of somatic slippage was predicted. Extensions of the approach

taken in the chapter should be used to accurately determine somatic MS slippage rates, and could be used to evaluate the efficacy of panels of MS markers used for the detection of genetic instability.

In this thesis, it has been shown how mathematical modelling, closely coupled with laboratory investigation, can be used to produce and test novel, complex hypotheses about the processes of clonal expansion and tumorigenesis in the human colon. I hope that mathematical models, as a formalised description of biological theory, will assume a central role in future attempts to decipher complex biological processes, particularly tumorigenesis.

Bibliography

ALBERICI, P., E. DE PATER, J. CARDOSO, M. BEVELANDER, L. MOLENAAR, J. JONKERS, and R. FODDE, 2007 Aneuploidy arises at early stages of Apc-driven intestinal tumorigenesis and pinpoints conserved chromosomal loci of allelic imbalance between mouse and human. Am J Pathol **170**: 377–87.

ALBUQUERQUE, C., C. BREUKEL, R. VAN DER LUIJT, P. FIDALGO, P. LAGE, F. J. M. SLORS, C. N. LEITÃO, R. FODDE, and R. SMITS, 2002 The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. Hum Mol Genet **11**: 1549–60.

ANDREYEV, H. J., A. R. NORMAN, D. CUNNINGHAM, J. OATES, B. R. DIX, B. J. IACOPETTA, J. YOUNG, T. WALSH, R. WARD, N. HAWKINS, M. BERANEK, P. JANDIK, R. BENAMOUZIG, E. JULLIAN, P. LAURENT-PUIG, S. OLSCHWANG, O. MULLER, I. HOFFMANN, H. M. RABES, C. ZIETZ, C. TROUNGOS, C. VALAVANIS, S. T. YUEN, J. W. HO, C. T. CROKE, D. P. O'DONOGHUE, W. GIARETTI, A. RAPALLO, A. RUSSO, V. BAZAN, M. TANAKA, K. OMURA, T. AZUMA, T. OHKUSA, T. FUJIMORI, Y. ONO, M. PAULY, C. FABER, R. GLAESENER, A. F. DE GOEIJ, J. W. ARENDS, S. N. ANDERSEN, T. LÖVIG, J. BREIVIK, G. GAUDERNACK, O. P. CLAUSEN, P. D. D. ANGELIS, G. I. MELING, T. O. ROGNUM, R. SMITH, H. S. GOH, A. FONT, R. ROSELL, X. F. SUN, H. ZHANG, J. BENHATTAR, L. LOSI, J. Q. LEE, S. T. WANG, P. A. CLARKE, S. BELL, P. QUIRKE, V. J. BUBB, J. PIRIS, N. R. CRUICKSHANK, D. MORTON, J. C. FOX, F. AL-MULLA, N. LEES, C. N. HALL, D. SNARY, K. WILKINSON, D. DILLON, J. COSTA, V. E. PRICOLO, S. D. FINKELSTEIN, J. S. THEBO, A. J. SENAGORE, S. A. HALTER, S. WADLER, S. MALIK, K. KRTOLICA, and N. UROSEVIC, 2001 Kirsten ras mutations in patients with colorectal cancer: the

'RASCAL II' study. Br J Cancer **85**: 692–6.

ANDREYEV, H. J., A. R. NORMAN, D. CUNNINGHAM, J. R. OATES, and P. A. CLARKE, 1998 Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. J Natl Cancer Inst **90**: 675–84.

AOKI, T., S. TAKEDA, A. YANAGISAWA, Y. KATO, Y. AJIOKA, H. WATANABE, S. KUDO, and Y. NAKAMURA, 1994 APC and p53 mutations in de novo colorectal adenocarcinomas. Hum Mutat **3**: 342–6.

ARMITAGE, P. and R. DOLL, 1954 The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer **8**: 1–12.

BAKER, S. J., E. R. FEARON, J. M. NIGRO, S. R. HAMILTON, A. C. PREISINGER, J. M. JESSUP, P. VANTUINEN, D. H. LEDBETTER, D. F. BARKER, Y. NAKAMURA, R. WHITE, and B. VOGELSTEIN, 1989 Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science **244**: 217–21.

BAMFORD, S., E. DAWSON, S. FORBES, J. CLEMENTS, R. PETTETT, A. DOGAN, A. FLANAGAN, J. TEAGUE, P. A. FUTREAL, M. R. STRATTON, and R. WOOSTER, 2004 The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer **91**: 355–8.

BARKER, N., J. H. VAN ES, J. KUIPERS, P. KUJALA, M. VAN DEN BORN, M. COZIJNSEN, A. HAEGEBARTH, J. KORVING, H. BEGTHEL, P. J. PETERS, and H. CLEVERS, 2007 Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature **449**: 1003–7.

BEAUMONT, M. A., W. ZHANG, and D. J. BALDING, 2002 Approximate Bayesian computation in population genetics. Genetics **162**: 2025–35.

BEERENWINKEL, N., T. ANTAL, D. DINGLI, A. TRAULSEN, K. W. KINZLER, V. E. VELCULESCU, B. VOGELSTEIN, and M. A. NOWAK, 2007 Genetic progression and the waiting time to cancer. PLoS Comput Biol **3**: e225.

BHATNAGAR, B., C. SHARMA, S. GUPTA, M. MATHUR, and D. REDDY, 2004 Study on the anatomical dimensions of the human sigmoid colon. Clin. Anat. **17**: 236–243.

BHATTACHARYYA, N. P., A. SKANDALIS, A. GANESH, J. GRODEN, and M. MEUTH, 1994 Mutator phenotypes in human colorectal carcinoma cell lines. Proc Natl Acad Sci USA **91**: 6319–23.

BJERKNES, M., 1986 A test of the stochastic theory of stem cell differentiation. Biophys J **49**: 1223–7.

BJERKNES, M. and H. CHENG, 1999 Clonal analysis of mouse intestinal epithelial progenitors. Gastroenterology **116**: 7–14.

BJERKNES, M., H. CHENG, K. HAY, and S. GALLINGER, 1997 APC mutation and the crypt cycle in murine and human intestine. Am J Pathol **150**: 833–9.

BODMER, W. F., C. J. BAILEY, J. BODMER, H. J. BUSSEY, A. ELLIS, P. GORMAN, F. C. LUCIBELLO, V. A. MURDAY, S. H. RIDER, and P. SCAMBLER, 1987 Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature **328**: 614–6.

BOLAND, C. R., S. N. THIBODEAU, S. R. HAMILTON, D. SIDRANSKY, J. R. ESHLEMAN, R. W. BURT, S. J. MELTZER, M. A. RODRIGUEZ-BIGAS, R. FODDE, G. N. RANZANI, and S. SRIVASTAVA, 1998 A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res **58**: 5248–57.

BOMAN, B. M., J. Z. FIELDS, O. BONHAM-CARTER, and O. A. RUNQUIST, 2001 Computer modeling implicates stem cell overproduction in colon cancer initiation. Cancer Res **61**: 8408–11.

BOYLE, M., C. WONG, M. ROCHA, and D. L. JONES, 2007 Decline in self-renewal factors contributes to aging of the stem cell niche in the drosophila testis. Cell Stem Cell **1**: 470–8.

BRAAKHUIS, B. J. M., M. P. TABOR, J. A. KUMMER, C. R. LEEMANS, and R. H. BRAKENHOFF, 2003 A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. Cancer Res **63**: 1727–30.

BRITTAN, M. and N. A. WRIGHT, 2004a The gastrointestinal stem cell. Cell Prolif **37**: 35–53.

BRITTAN, M. and N. A. WRIGHT, 2004b Stem cell in gastrointestinal structure and neoplastic development. Gut **53**: 899–910.

CAIRNS, J., 2002 Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. Proc Natl Acad Sci USA **99**: 10567–70.

CALABRESE, P., S. TAVARÉ, and D. SHIBATA, 2004 Pretumor progression: clonal evolution of human stem cell populations. Am J Pathol **164**: 1337–46.

CAMPBELL, F., M. A. APPLETON, C. E. FULLER, M. P. GREEFF, J. HALLGRIMSSON, R. KATOH, O. L. NG, A. SATIR, G. T. WILLIAMS, and E. D. WILLIAMS, 1994a Racial variation in the O-acetylation phenotype of human colonic mucosa. J Pathol **174**: 169–74.

CAMPBELL, F., C. E. FULLER, G. T. WILLIAMS, and E. D. WILLIAMS, 1994b Human colonic stem cell mutation frequency with and without irradiation. J Pathol **174**: 175–82.

CAMPBELL, F., G. T. WILLIAMS, M. A. APPLETON, M. F. DIXON, M. HARRIS, and E. D. WILLIAMS, 1996 Post-irradiation somatic mutation and clonal stabilisation time in the human colon. Gut **39**: 569–73.

CANCER RESEARCH UK, CancerStats: Incidence UK. http://publications.cancerresearchuk.org/WebRoot/crukstoredb/CRUK_PDFs/CSINC08.pdf.

CASTAGNOLA, P. and W. GIARETTI, 2005 Mutant KRAS, chromosomal instability and prognosis in colorectal cancer. Biochim Biophys Acta **1756**: 115–25.

CHEN, R., P. S. RABINOVITCH, D. A. CRISPIN, M. J. EMOND, M. P. BRONNER, and T. A. BRENTNALL, 2005 The initiation of colon cancer in a chronic inflammatory setting. Carcinogenesis **26**: 1513–9.

CHENG, H., M. BJERKNES, J. AMAR, and G. GARDINER, 1986a Crypt production in normal and diseased human colonic epithelium. Anat Rec **216**: 44–8.

CHENG, H., M. BJERKNES, J. AMAR, and G. GARDINER, 1986b Crypt production in normal and diseased human colonic epithelium. Anat Rec **216**: 44–8.

CHENG, H., C. MCCULLOCH, and M. BJERKNES, 1986c Effects of 30% intestinal resection on whole population cell kinetics of mouse intestinal epithelium. Anat Rec **215**: 35–41.

CHENG, L., K. ARAKI, Y. FURUYA, T. MATSUOKA, K. MASHIMA, M. KOBAYASHI, and K. MATSUURA, 2000 Morphological study of the regeneration mechanism of acetic acid-injured colon crypts in the rat. Medical electron microscopy : official journal of the Clinical Electron Microscopy Society of Japan **33**: 165–71.

CHU, M. W., K. D. SIEGMUND, C. L. ECKSTAM, J. Y. KIM, A. S. YANG, G. C. KANEL, S. TAVARÉ, and D. SHIBATA, 2007 Lack of increases in methylation at three CpG-rich genomic loci in non-mitotic adult tissues during aging. BMC Med Genet **8**: 50.

COLLER, H., K. KHRAPKO, N. BODYAK, and E. NEKHAEVA, 2001 High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without .... Nature Genetics .

CRABTREE, M., O. M. SIEBER, L. LIPTON, S. V. HODGSON, H. LAMLUM, H. J. W. THOMAS, K. NEALE, R. K. S. PHILLIPS, K. HEINIMANN, and I. P. M. TOMLINSON, 2003 Refining the relation between 'first hits' and 'second hits' at the APC locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. Oncogene **22**: 4257–65.

CRABTREE, M. D., I. P. M. TOMLINSON, S. V. HODGSON, K. NEALE, R. K. S. PHILLIPS, and R. S. HOULSTON, 2002 Explaining variation in familial adenomatous polyposis: relationship between genotype and phenotype and evidence for modifier genes. Gut **51**: 420–3.

CROSNIER, C., D. STAMATAKI, and J. LEWIS, 2006 Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. Nat Rev Genet **7**: 349–59.

DE ABAJO, A. S., M. DE LA HOYA, M. VAN PUIJENBROEK, A. TOSAR, J. A. LÓPEZ-ASENJO, E. DÍAZ-RUBIO, H. MORREAU, and T. CALDES, 2007 Molecular

analysis of colorectal cancer tumors from patients with mismatch repair proficient hereditary nonpolyposis colorectal cancer suggests novel carcinogenic pathways. Clin Cancer Res **13**: 5729–35.

DE LA CHAPELLE, A., 2003 Microsatellite instability. N Engl J Med **349**: 209–10.

DE LA CHAPELLE, A., 2004 Genetic predisposition to colorectal cancer. Nature Reviews Cancer **4**: 769–80.

DEN DUNNEN, J. and S. ANTONARAKIS, 2001 Nomenclature for the description of human sequence variations. Hum Genet .

DEUTER, R., J. LINZ, S. PIETSCH, G. WINDE, S. HENTSCH, and O. MULLER, 1996 DNA alterations in sporadic colorectal tumors do not correlate with tumor staging diagnosed by the TNM system. Cancer Lett **109**: 161–9.

DOMINGO, E., E. ESPÍN, M. ARMENGOL, C. OLIVEIRA, M. PINTO, A. DUVAL, C. BRENNETOT, R. SERUCA, R. HAMELIN, H. YAMAMOTO, and S. SCHWARTZ, 2004 Activated BRAF targets proximal colon tumors with mismatch repair deficiency and MLH1 inactivation. Genes Chromosomes Cancer **39**: 138–42.

EDLER, L. and A. KOPP-SCHNEIDER, 2005 Origins of the mutational origin of cancer. International Journal of Epidemiology **34**: 1168–70.

EDWARDS, C. M. and S. J. CHAPMAN, 2007 Biomechanical Modelling of Colorectal Crypt Budding and Fission. Bull. Math. Biol. **69**: 1927–1942.

ELMORE, E., T. KAKUNAGA, and J. C. BARRETT, 1983 Comparison of spontaneous mutation rates of normal and chemically transformed human skin fibroblasts. Cancer Res **43**: 1650–5.

ELSON, J. L., D. C. SAMUELS, D. M. TURNBULL, and P. F. CHINNERY, 2001 Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. Am J Hum Genet **68**: 802–6.

ESPINA, V., J. WULFKUHLE, V. CALVERT, and A. VANMETER, 2006 Laser-capture microdissection. Nature Protocols **1**: 586–603.

ETIENNE-MANNEVILLE, S. and A. HALL, 2003 Cdc42 regulates GSK-3beta and adenomatous polyposis coli to control cell polarity. Nature **421**: 753–6.

FARR, C. J., C. J. MARSHALL, D. J. EASTY, N. A. WRIGHT, S. C. POWELL, and C. PARASKEVA, 1988 A study of ras gene mutations in colonic adenomas from familial polyposis coli patients. Oncogene **3**: 673–8.

FAUX, M. C., J. L. ROSS, C. MEEKER, T. JOHNS, H. JI, R. J. SIMPSON, M. J. LAYTON, and A. W. BURGESS, 2004 Restoration of full-length adenomatous polyposis coli (APC) protein in a colon cancer cell line enhances cell adhesion. J Cell Sci **117**: 427–39.

FEARON, E. R. and B. VOGELSTEIN, 1990 A genetic model for colorectal tumorigenesis. Cell **61**: 759–67.

FEVR, T., S. ROBINE, D. LOUVARD, and J. HUELSKEN, 2007 Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. Mol Cell Biol **27**: 7551–9.

FISHEL, R., M. K. LESCOE, M. R. RAO, N. G. COPELAND, N. A. JENKINS, J. GARBER, M. KANE, and R. KOLODNER, 1993 The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell **75**: 1027–38.

FODDE, R., J. KUIPERS, C. ROSENBERG, R. SMITS, M. KIELMAN, C. GASPAR, J. H. VAN ES, C. BREUKEL, J. WIEGANT, R. H. GILES, and H. CLEVERS, 2001 Mutations in the APC tumour suppressor gene cause chromosomal instability. Nat Cell Biol **3**: 433–8.

FORBES, S., J. CLEMENTS, E. DAWSON, S. BAMFORD, T. WEBB, A. DOGAN, A. FLANAGAN, J. TEAGUE, R. WOOSTER, P. A. FUTREAL, and M. R. STRATTON, 2006 COSMIC 2005. Br J Cancer **94**: 318–22.

FRAYLING, I. M., N. E. BECK, M. ILYAS, I. DOVE-EDWIN, P. GOODMAN, K. PACK, J. A. BELL, C. B. WILLIAMS, S. V. HODGSON, H. J. THOMAS, I. C. TALBOT, W. F. BODMER, and I. P. TOMLINSON, 1998 The APC variants I1307K

and E1317Q are associated with colorectal tumors, but not always with a family history. Proc Natl Acad Sci USA **95**: 10722–7.

GÖNCZY, P. and S. DINARDO, 1996 The germ line regulates somatic cyst cell proliferation and fate during drosophila spermatogenesis. Development **122**: 2437–47.

GONZÁLEZ-GARCÍA, I., V. MORENO, M. NAVARRO, J. MARTÍ-RAGUÉ, E. MARCUELLO, C. BENASCO, O. CAMPOS, G. CAPELLÀ, and M. A. PEINADO, 2000 Standardized approach for microsatellite instability detection in colorectal carcinomas. J Natl Cancer Inst **92**: 544–9.

GONZÁLEZ-GARCÍA, I., R. V. SOLÉ, and J. COSTA, 2002 Metapopulation dynamics and spatial heterogeneity in cancer. Proc Natl Acad Sci USA **99**: 13085–9.

GRADY, W. M., 2004 Genomic instability and colon cancer. Cancer Metastasis Rev **23**: 11–27.

GRAHAM, T., S. HALFORD, K. M. PAGE, and I. P. M. TOMLINSON, 2008 Most low-level microsatellite instability in colorectal cancers can be explained without an elevated slippage rate. J Pathol **215**: 204–10.

GREAVES, L. C., S. L. PRESTON, P. J. TADROUS, R. W. TAYLOR, M. J. BARRON, D. OUKRIF, S. J. LEEDHAM, M. DEHERAGODA, P. SASIENI, M. R. NOVELLI, J. A. Z. JANKOWSKI, D. M. TURNBULL, N. A. WRIGHT, and S. A. C. MCDONALD, 2006 Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. Proc Natl Acad Sci USA **103**: 714–9.

GREGORIEFF, A., D. PINTO, H. BEGTHEL, O. DESTRÉE, M. KIELMAN, and H. CLEVERS, 2005 Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology **129**: 626–38.

GRISHAM, M. B., 1994 Oxidants and free radicals in inflammatory bowel disease. Lancet **344**: 859–61.

GRODEN, J., A. THLIVERIS, W. SAMOWITZ, M. CARLSON, L. GELBERT, H. ALBERTSEN, G. JOSLYN, J. STEVENS, L. SPIRIO, and M. ROBERTSON, 1991

Identification and characterization of the familial adenomatous polyposis coli gene. Cell **66**: 589–600.

HALFORD, S., 2005 *Microsatellite instability in colorectal cancer*. PhD, University of London.

HALFORD, S., A. ROWAN, E. SAWYER, I. TALBOT, and I. TOMLINSON, 2005 O(6)-methylguanine methyltransferase in colorectal cancers: detection of mutations, loss of expression, and weak association with G:C>A:T transitions. Gut **54**: 797–802.

HALFORD, S., P. SASIENI, A. ROWAN, H. WASAN, W. BODMER, I. TALBOT, N. HAWKINS, R. WARD, and I. TOMLINSON, 2002 Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait. Cancer Res **62**: 53–7.

HALFORD, S. E. R., A. J. ROWAN, L. LIPTON, O. M. SIEBER, K. PACK, H. J. W. THOMAS, S. V. HODGSON, W. F. BODMER, and I. P. M. TOMLINSON, 2003 Germline mutations but not somatic changes at the MYH locus contribute to the pathogenesis of unselected colorectal cancers. Am J Pathol **162**: 1545–8.

HANAHAN, D. and R. A. WEINBERG, 2000 The hallmarks of cancer. Cell **100**: 57–70.

HANFORD, M. G., B. C. RUSHTON, L. C. GOWEN, and R. A. FARBER, 1998 Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair. Oncogene **16**: 2389–93.

HAYASHI, K., 1991 PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl **1**: 34–8.

HOEIJMAKERS, J. H., 2001 Genome maintenance mechanisms for preventing cancer. Nature **411**: 366–74.

HORNSBY, C., K. M. PAGE, and I. TOMLINSON, 2008 The in vivo rate of somatic adenomatous polyposis coli mutation. Am J Pathol **172**: 1062–8.

HOULSTON, R. S., 2001 What we could do now: molecular pathology of colorectal cancer. MP, Mol Pathol **54**: 206–14.

HOUNNOU, G., C. DESTRIEUX, J. DESMÉ, P. BERTRAND, and S. VELUT, 2002 Anatomical study of the length of the human intestine. Surgical and radiologic anatomy : SRA **24**: 290–4.

HUMPHRIES, A. and N. A. WRIGHT, 2008 Colonic crypt organization and tumorigenesis. Nature Reviews Cancer **8**: 415–24.

ILLUMINA, DNA Sequencing with Solexa ®Technology. http://www.illumina.com/downloads/SS_DNAsequencing.pdf.

IONOV, Y., M. A. PEINADO, S. MALKHOSYAN, D. SHIBATA, and M. PERUCHO, 1993 Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature **363**: 558–61.

IYER, R. R., A. PLUCIENNIK, V. BURDETT, and P. L. MODRICH, 2006 DNA mismatch repair: functions and mechanisms. Chem Rev **106**: 302–23.

JANSSEN, K.-P., M. ABAL, M. ABALA, F. E. MARJOU, D. LOUVARD, and S. ROBINE, 2005 Mouse models of K-ras-initiated carcinogenesis. Biochim Biophys Acta **1756**: 145–54.

JASS, J. R., 2007 Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology **50**: 113–30.

JASS, J. R., K. G. BIDEN, M. C. CUMMINGS, L. A. SIMMS, M. WALSH, E. SCHOCH, S. J. MELTZER, C. WRIGHT, J. SEARLE, J. YOUNG, and B. A. LEGGETT, 1999 Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. J Clin Pathol **52**: 455–60.

JEN, J., S. M. POWELL, N. PAPADOPOULOS, K. J. SMITH, S. R. HAMILTON, B. VOGELSTEIN, and K. W. KINZLER, 1994 Molecular determinants of dysplasia in colorectal lesions. Cancer Res **54**: 5523–6.

JEON, C.-H., H.-I. LEE, I.-H. SHIN, and J.-W. PARK, 2008 Genetic alterations of APC, K-ras, p53, MSI, and MAGE in Korean colorectal cancer patients. International journal of colorectal disease **23**: 29–35.

JOHNSON, L., D. GREENBAUM, K. CICHOWSKI, K. MERCER, E. MURPHY, E. SCHMITT, R. T. BRONSON, H. UMANOFF, W. EDELMANN, R. KUCHERLAPATI, and T. JACKS, 1997 K-ras is an essential gene in the mouse with partial functional overlap with N-ras. Genes Dev **11**: 2468–81.

JOHNSON, L., K. MERCER, D. GREENBAUM, R. T. BRONSON, D. CROWLEY, D. A. TUVESON, and T. JACKS, 2001 Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature **410**: 1111–6.

JONES, A., R. MITTER, R. SPRINGALL, T. GRAHAM, E. WINTER, C. GILLETT, A. HANBY, I. TOMLINSON, and E. SAWYER, 2008a A comprehensive genetic profile of phyllodes tumours of the breast detects important mutations, intra-tumoral genetic heterogeneity and new genetic changes on recurrence. J Pathol **214**: 533–544.

JONES, A., C. THIRLWELL, K. HOWARTH, T. GRAHAM, W. CHAMBERS, S. SEGDITSAS, K. PAGE, R. PHILLIPS, H. THOMAS, O. SIEBER, E. SAWYER, and I. TOMLINSON, 2007 Analysis of copy number changes suggests chromosomal instability in a minority of large colorectal adenomas. J Pathol **213**: 249–56.

JONES, S., W.-D. CHEN, G. PARMIGIANI, F. DIEHL, N. BEERENWINKEL, T. ANTAL, A. TRAULSEN, M. A. NOWAK, C. SIEGEL, V. E. VELCULESCU, K. W. KINZLER, B. VOGELSTEIN, J. WILLIS, and S. D. MARKOWITZ, 2008b Comparative lesion sequencing provides insights into tumor evolution. Proc Natl Acad Sci USA **105**: 4283–8.

KANAOKA, S., H. HANAI, E. KANEKO, I. KINO, S. BABA, and M. FUJITA, 1996 A very large villous adenoma with an adjacent cancer of the rectum: an informative case for testing the proposed molecular basis of colorectal tumorigenesis. Jpn J Clin Oncol **26**: 384–90.

KANE, M. F., M. LODA, G. M. GAIDA, J. LIPMAN, R. MISHRA, H. GOLDMAN, J. M. JESSUP, and R. KOLODNER, 1997 Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res **57**: 808–11.

KANNAN, K. and S. JAIN, 2000 Oxidative stress and apoptosis. Pathophysiology **7**: 153–163.

KAWASAKI, T., M. OHNISHI, K. NOSHO, Y. SUEMOTO, G. J. KIRKNER, J. A. MEYERHARDT, C. S. FUCHS, and S. OGINO, 2008 CpG island methylator phenotype-low (CIMP-low) colorectal cancer shows not only few methylated CIMP-high-specific CpG islands, but also low-level methylation at individual loci. Mod Pathol **21**: 245–55.

KETS, C. M., N. HOOGERBRUGGE, D. BODMER, R. WILLEMS, H. G. BRUNNER, J. H. J. M. VAN KRIEKEN, and M. J. L. LIGTENBERG, 2006 Unfavorable pathological characteristics in familial colorectal cancer with low-level microsatellite instability. Mod Pathol **19**: 1624–30.

KIGER, A. A., D. L. JONES, C. SCHULZ, M. B. ROGERS, and M. T. FULLER, 2001 Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. Science **294**: 2542–5.

KIM, I.-J., J.-L. KU, H. C. KANG, J.-H. PARK, K.-A. YOON, Y. SHIN, H.-W. PARK, S. G. JANG, S.-K. LIM, S. Y. HAN, Y.-K. SHIN, M. R. LEE, S.-Y. JEONG, H.-R. SHIN, J. S. LEE, W.-H. KIM, and J.-G. PARK, 2004a Mutational analysis of OGG1, MYH, MTH1 in FAP, HNPCC and sporadic colorectal cancer patients: R154H OGG1 polymorphism is associated with sporadic colorectal cancer patients. Hum Genet **115**: 498–503.

KIM, J. Y., R. W. BEART, and D. SHIBATA, 2005a Stability of colon stem cell methylation after neo-adjuvant therapy in a patient with attenuated familial adenomatous polyposis. BMC gastroenterology **5**: 19.

KIM, J. Y., R. W. BEART, and D. SHIBATA, 2005b Stability of colon stem cell methylation after neo-adjuvant therapy in a patient with attenuated familial adenomatous polyposis. BMC gastroenterology **5**: 19.

KIM, J. Y., K. D. SIEGMUND, S. TAVARÉ, and D. SHIBATA, 2005c Age-related human small intestine methylation: evidence for stem cell niches. BMC medicine **3**: 10.

KIM, J. Y., S. TAVARÉ, and D. SHIBATA, 2005d Counting human somatic cell replications: methylation mirrors endometrial stem cell divisions. Proc Natl Acad Sci USA **102**: 17739–44.

KIM, J. Y., S. TAVARÉ, and D. SHIBATA, 2006 Human hair genealogies and stem cell latency. BMC Biol **4**: 2.

KIM, K.-M., P. CALABRESE, S. TAVARÉ, and D. SHIBATA, 2004b Enhanced stem cell survival in familial adenomatous polyposis. Am J Pathol **164**: 1369–77.

KIM, K.-M., P. CALABRESE, S. TAVARÉ, and D. SHIBATA, 2004c Enhanced stem cell survival in familial adenomatous polyposis. Am J Pathol **164**: 1369–77.

KIM, K.-M. and D. SHIBATA, 2004 Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. BMC gastroenterology **4**: 8.

KINZLER, K. W. and B. VOGELSTEIN, 1996 Lessons from hereditary colorectal cancer. Cell **87**: 159–70.

KINZLER, K. W. and B. VOGELSTEIN, 1997 Cancer-susceptibility genes. Gatekeepers and caretakers. Nature **386**: 761, 763.

KIRKWOOD, B. and J. STERNE, 2003 *Essential Medical Statistics.*. Wiley Blackwell, second edition.

KLAUS, A. and W. BIRCHMEIER, 2008 Wnt signalling and its impact on development and cancer. Nature Reviews Cancer **8**: 387–98.

KNUDSON, A. G., 1971 Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA **68**: 820–3.

KNUDSON, A. G., 2001 Two genetic hits (more or less) to cancer. Nature Reviews Cancer **1**: 157–62.

KOERA, K., K. NAKAMURA, K. NAKAO, J. MIYOSHI, K. TOYOSHIMA, T. HATTA, H. OTANI, A. AIBA, and M. KATSUKI, 1997 K-ras is essential for the development of the mouse embryo. Oncogene **15**: 1151–9.

KOMAROVA, N. L., 2005 Cancer, aging and the optimal tissue design. Seminars in Cancer Biology **15**: 494–505.

KOMAROVA, N. L., C. LENGAUER, B. VOGELSTEIN, and M. A. NOWAK, 2002 Dynamics of genetic instability in sporadic and familial colorectal cancer. Cancer Biol Ther **1**: 685–92.

KORINEK, V., N. BARKER, P. MOERER, E. VAN DONSELAAR, G. HULS, P. J. PETERS, and H. CLEVERS, 1998 Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet **19**: 379–83.

KOSINSKI, C., V. S. W. LI, A. S. Y. CHAN, J. ZHANG, C. HO, W. Y. TSUI, T. L. CHAN, R. C. MIFFLIN, D. W. POWELL, S. T. YUEN, S. Y. LEUNG, and X. CHEN, 2007 Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. Proc Natl Acad Sci USA **104**: 15418–23.

KRUGLYAK, S., R. T. DURRETT, M. D. SCHUG, and C. F. AQUADRO, 1998 Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA **95**: 10774–8.

LAI, Y. and F. SUN, 2003 The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol Evol **20**: 2123–31.

LAIHO, P., V. LAUNONEN, P. LAHERMO, M. ESTELLER, M. GUO, J. G. HERMAN, J.-P. MECKLIN, H. JÄRVINEN, P. SISTONEN, K.-M. KIM, D. SHIBATA, R. S. HOULSTON, and L. A. AALTONEN, 2002 Low-level microsatellite instability in most colorectal carcinomas. Cancer Res **62**: 1166–70.

LAMLUM, H., M. ILYAS, A. ROWAN, S. CLARK, V. JOHNSON, J. BELL, I. FRAYLING, J. EFSTATHIOU, K. PACK, S. PAYNE, R. ROYLANCE, P. GORMAN, D. SHEER, K. NEALE, R. PHILLIPS, I. TALBOT, W. BODMER, and I. TOMLINSON, 1999 The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. Nature Medicine **5**: 1071–5.

LAMLUM, H., N. A. TASSAN, E. JAEGER, I. FRAYLING, O. SIEBER, F. B. REZA, M. ECKERT, A. ROWAN, E. BARCLAY, W. ATKIN, C. WILLIAMS, J. GILBERT, J. CHEADLE, J. BELL, R. HOULSTON, W. BODMER, J. SAMPSON, and I. TOMLINSON, 2000 Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. Hum Mol Genet **9**: 2215–21.

LEEDHAM, S. J., M. BRITTAN, S. A. C. MCDONALD, and N. A. WRIGHT, 2005 Intestinal stem cells. J Cell Mol Med **9**: 11–24.

LEPPERT, M., M. DOBBS, P. SCAMBLER, P. O'CONNELL, Y. NAKAMURA, D. STAUFFER, S. WOODWARD, R. BURT, J. HUGHES, and E. GARDNER, 1987 The gene for familial polyposis coli maps to the long arm of chromosome 5. Science **238**: 1411–3.

LESSARD, J. and G. SAUVAGEAU, 2003 Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature **423**: 255–60.

LI, J., Y. MIZUKAMI, X. ZHANG, W.-S. JO, and D. C. CHUNG, 2005 Oncogenic K-ras stimulates Wnt signaling in colon cancer through inhibition of GSK-3beta. Gastroenterology **128**: 1907–18.

LILLEBERG, S. L., J. DUROCHER, C. SANDERS, K. WALTERS, and K. CULVER, 2004 High sensitivity scanning of colorectal tumors and matched plasma DNA for mutations in APC, TP53, K-RAS, and BRAF genes with a novel DHPLC fluorescence detection platform. Ann N Y Acad Sci **1022**: 250–6.

LOEFFLER, M., A. BIRKE, D. WINTON, and C. POTTEN, 1993 Somatic mutation, monoclonality and stochastic models of stem cell organization in the intestinal crypt. J Theor Biol **160**: 471–91.

LOEFFLER, M., T. BRATKE, U. PAULUS, Y. Q. LI, and C. S. POTTEN, 1997 Clonality and life cycles of intestinal crypts explained by a state dependent stochastic model of epithelial stem cell organization. J Theor Biol **186**: 41–54.

LUEBECK, E. G. and S. H. MOOLGAVKAR, 2002 Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci USA **99**: 15095–100.

MAHMOUD, N. N., S. K. BOOLBOL, R. T. BILINSKI, C. MARTUCCI, A. CHADBURN, and M. M. BERTAGNOLLI, 1997 Apc gene mutation is associated with a dominant-negative effect upon intestinal cell migration. Cancer Res **57**: 5045–50.

MCCOMBS, R., 1930 A hypothesis on the causation of cancer. Science **72**: 423–424.

MCDONALD, S. A. C., S. L. PRESTON, L. C. GREAVES, S. J. LEEDHAM, M. A. LOVELL, J. A. Z. JANKOWSKI, D. M. TURNBULL, and N. A. WRIGHT, 2006a Clonal expansion in the human gut: mitochondrial DNA mutations show us the way. Cell Cycle **5**: 808–11.

MCDONALD, S. A. C., S. L. PRESTON, M. J. LOVELL, N. A. WRIGHT, and J. A. Z. JANKOWSKI, 2006b Mechanisms of disease: from stem cells to colorectal cancer. Nature clinical practice Gastroenterology & hepatology **3**: 267–74.

MCLELLAN, E. A., R. A. OWEN, K. A. STEPNIEWSKA, J. P. SHEFFIELD, and N. R. LEMOINE, 1993 High frequency of K-ras mutations in sporadic colorectal adenomas. Gut **34**: 392–6.

MEINEKE, F. A., C. S. POTTEN, and M. LOEFFLER, 2001 Cell migration and organization in the intestinal crypt using a lattice-free model. Cell Prolif **34**: 253–66.

MIYAKI, M., T. IIJIMA, R. ISHII, Y. KITA, M. KOIKE, T. KUROKI, and T. MORI, 2002 Increased frequency of p53 mutation in sporadic colorectal cancer from cigarette smokers. Jpn J Clin Oncol **32**: 196–201.

MIYAKI, M., T. IIJIMA, M. OHUE, Y. KITA, T. HISHIMA, T. KUROKI, T. IWAMA, and T. MORI, 2003 A novel case with germline p53 gene mutation having concurrent multiple primary colon tumours. Gut **52**: 304–6.

MIYAKI, M., J. NISHIO, M. KONISHI, R. KIKUCHI-YANOSHITA, K. TANAKA, M. MURAOKA, M. NAGATO, J. M. CHONG, M. KOIKE, T. TERADA, Y. KAWAHARA, A. FUKUTOME, J. TOMIYAMA, Y. CHUGANJI, M. MOMOI, and J. UTSUNOMIYA, 1997 Drastic genetic instability of tumors and normal tissues in Turcot syndrome. Oncogene **15**: 2877–81.

MIYOSHI, Y., H. NAGASE, H. ANDO, A. HORII, S. ICHII, S. NAKATSURU, T. AOKI, Y. MIKI, T. MORI, and Y. NAKAMURA, 1992 Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet **1**: 229–33.

MOLOFSKY, A. V., R. PARDAL, T. IWASHITA, I.-K. PARK, M. F. CLARKE, and S. J. MORRISON, 2003 Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature **425**: 962–7.

MOOLGAVKAR, S. H. and E. G. LUEBECK, 2003 Multistage carcinogenesis and the incidence of human cancer. Genes Chromosomes Cancer **38**: 302–6.

MOSER, A. R., H. C. PITOT, and W. F. DOVE, 1990 A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science **247**: 322–4.

NICOLAS, P., K.-M. KIM, D. SHIBATA, and S. TAVARÉ, 2007 The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput Biol **3**: e28.

NOVELLI, M., A. COSSU, D. OUKRIF, A. QUAGLIA, S. LAKHANI, R. POULSOM, P. SASIENI, P. CARTA, M. CONTINI, A. PASCA, G. PALMIERI, W. BODMER, F. TANDA, and N. WRIGHT, 2003 X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. Proc Natl Acad Sci USA **100**: 3311–4.

NOVELLI, M. R., J. A. WILLIAMSON, I. P. TOMLINSON, G. ELIA, S. V. HODGSON, I. C. TALBOT, W. F. BODMER, and N. A. WRIGHT, 1996 Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. Science **272**: 1187–90.

NOWAK, M. A., N. L. KOMAROVA, A. SENGUPTA, P. V. JALLEPALLI, I.-M. SHIH, B. VOGELSTEIN, and C. LENGAUER, 2002 The role of chromosomal instability in tumor initiation. Proc Natl Acad Sci USA **99**: 16226–31.

NUCCI, M. R., C. R. ROBINSON, P. LONGO, P. CAMPBELL, and S. R. HAMILTON, 1997 Phenotypic and genotypic characteristics of aberrant crypt foci in human colorectal mucosa. Human Pathology **28**: 1396–407.

NUGENT, K. P., R. K. PHILLIPS, S. V. HODGSON, S. COTTRELL, J. SMITH-RAVIN, K. PACK, and W. F. BODMER, 1994 Phenotypic expression in familial adenomatous

polyposis: partial prediction by mutation analysis. Gut **35**: 1622–3.

OGINO, S., T. KAWASAKI, G. J. KIRKNER, M. LODA, and C. S. FUCHS, 2006 CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. The Journal of molecular diagnostics : JMD **8**: 582–8.

OTORI, K., M. KONISHI, K. SUGIYAMA, and T. HASEBE, 1998 Infrequent somatic mutation of the adenomatous polyposis coli gene in aberrant crypt foci of human .... Cancer .

PAPADOPOULOS, N., N. C. NICOLAIDES, Y. F. WEI, S. M. RUBEN, K. C. CARTER, C. A. ROSEN, W. A. HASELTINE, R. D. FLEISCHMANN, C. M. FRASER, and M. D. ADAMS, 1994 Mutation of a mutL homolog in hereditary colon cancer. Science **263**: 1625–9.

PARK, K.-S., S. H. JEON, S.-E. KIM, Y.-Y. BAHK, S. L. HOLMEN, B. O. WILLIAMS, K.-C. CHUNG, Y.-J. SURH, and K.-Y. CHOI, 2006 APC inhibits ERK pathway activation and cellular proliferation induced by RAS. J Cell Sci **119**: 819–27.

PARSONS, R., G. M. LI, M. J. LONGLEY, W. H. FANG, N. PAPADOPOULOS, J. JEN, A. DE LA CHAPELLE, K. W. KINZLER, B. VOGELSTEIN, and P. MODRICH, 1993 Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell **75**: 1227–36.

PAVLICK, K. P., F. S. LAROUX, J. FUSELER, R. E. WOLF, L. GRAY, J. HOFFMAN, and M. B. GRISHAM, 2002 Role of reactive metabolites of oxygen and nitrogen in inflammatory bowel disease. Free Radic Biol Med **33**: 311–22.

PEPPER, J. W., K. SPROUFFSKE, and C. C. MALEY, 2007 Animal cell differentiation patterns suppress somatic evolution. PLoS Comput Biol **3**: e250.

PIERCE, G. B. and W. C. SPEERS, 1988 Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. Cancer Res **48**: 1996–2004.

POTTEN, C. S., C. BOOTH, and D. M. PRITCHARD, 1997 The intestinal epithelial stem cell: the mucosal governor. International journal of experimental pathology **78**: 219–43.

POTTEN, C. S. and M. LOEFFLER, 1990 Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development **110**: 1001–20.

POWELL, S. M., N. ZILZ, Y. BEAZER-BARCLAY, T. M. BRYAN, S. R. HAMILTON, S. N. THIBODEAU, B. VOGELSTEIN, and K. W. KINZLER, 1992 APC mutations occur early during colorectal tumorigenesis. Nature **359**: 235–7.

PRESTON, S. L., W.-M. WONG, A. O.-O. CHAN, R. POULSOM, R. JEFFERY, R. A. GOODLAD, N. MANDIR, G. ELIA, M. NOVELLI, W. F. BODMER, I. P. TOMLINSON, and N. A. WRIGHT, 2003 Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. Cancer Res **63**: 3819–25.

PRETLOW, T. P. and T. G. PRETLOW, 2005 Mutant KRAS in aberrant crypt foci (ACF): initiation of colorectal cancer? Biochim Biophys Acta **1756**: 83–96.

R DEVELOPMENT CORE TEAM, 2007 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

RACHMILEWITZ, D., J. S. STAMLER, D. BACHWICH, F. KARMELI, Z. ACKERMAN, and D. K. PODOLSKY, 1995 Enhanced colonic nitric oxide generation and nitric oxide synthase activity in ulcerative colitis and Crohn's disease. Gut **36**: 718–723.

RAJAGOPALAN, H., M. A. NOWAK, B. VOGELSTEIN, and C. LENGAUER, 2003 Opinion: The significance of unstable chromosomes in colorectal cancer. Nature Reviews Cancer **3**: 695–701.

RICE, J., 1994 *Mathematical Statistics and Data Analysis.*. Wadsworth, second edition.

ROWAN, A. J., H. LAMLUM, M. ILYAS, J. WHEELER, J. STRAUB, A. PAPADOPOULOU, D. BICKNELL, W. F. BODMER, and I. P. TOMLINSON, 2000 APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". Proc Natl Acad Sci USA **97**: 3352–7.

SANGIORGI, E. and M. R. CAPECCHI, 2008 Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet **40**: 915–20.

SCHNEIKERT, J. and J. BEHRENS, 2006 Truncated APC is required for cell proliferation and DNA replication. Int J Cancer **119**: 74–9.

SCHNEIKERT, J. and J. BEHRENS, 2007 The canonical Wnt signalling pathway and its APC partner in colon cancer development. Gut **56**: 417–25.

SEGDITSAS, S. and I. TOMLINSON, 2006 Colorectal cancer and genetic alterations in the Wnt pathway. Oncogene **25**: 7531–7.

SHEN, L., Y. KONDO, G. L. ROSNER, L. XIAO, N. S. HERNANDEZ, J. VILAYTHONG, P. S. HOULIHAN, R. S. KROUSE, A. R. PRASAD, J. G. EINSPAHR, J. BUCKMEIER, D. S. ALBERTS, S. R. HAMILTON, and J.-P. J. ISSA, 2005 MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst **97**: 1330–8.

SHIH, I. M., W. ZHOU, S. N. GOODMAN, C. LENGAUER, K. W. KINZLER, and B. VOGELSTEIN, 2001 Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. Cancer Res **61**: 818–22.

SIEBEN, N. L., N. T. TER HAAR, C. J. CORNELISSE, G. J. FLEUREN, and A. M. CLETON-JANSEN, 2000 PCR artifacts in LOH and MSI analysis of microdissected tumor cells. Human Pathology **31**: 1414–9.

SIEBER, O. M., K. HEINIMANN, P. GORMAN, H. LAMLUM, M. CRABTREE, C. A. SIMPSON, D. DAVIES, K. NEALE, S. V. HODGSON, R. R. ROYLANCE, R. K. S. PHILLIPS, W. F. BODMER, and I. P. M. TOMLINSON, 2002 Analysis of chromosomal instability in human colorectal adenomas with two mutational hits at APC. Proc Natl Acad Sci USA **99**: 16910–5.

SIEBER, O. M., K. HEINIMANN, and I. P. M. TOMLINSON, 2003 Genomic instability–the engine of tumorigenesis? Nature Reviews Cancer **3**: 701–8.

SINGER, I. I., D. W. KAWKA, S. SCOTT, J. R. WEIDNER, R. A. MUMFORD, T. E. RIEHL, and W. F. STENSON, 1996 Expression of inducible nitric oxide synthase and

nitrotyrosine in colonic epithelium in inflammatory bowel disease. Gastroenterology **111**: 871–85.

SJÖBLOM, T., S. JONES, L. D. WOOD, D. W. PARSONS, J. LIN, T. D. BARBER, D. MANDELKER, R. J. LEARY, J. PTAK, N. SILLIMAN, S. SZABO, P. BUCKHAULTS, C. FARRELL, P. MEEH, S. D. MARKOWITZ, J. WILLIS, D. DAWSON, J. K. V. WILLSON, A. F. GAZDAR, J. HARTIGAN, L. WU, C. LIU, G. PARMIGIANI, B. H. PARK, K. E. BACHMAN, N. PAPADOPOULOS, B. VOGELSTEIN, K. W. KINZLER, and V. E. VELCULESCU, 2006 The consensus coding sequences of human breast and colorectal cancers. Science **314**: 268–74.

SLAUGHTER, D. P., H. W. SOUTHWICK, and W. SMEJKAL, 1953 Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer **6**: 963–8.

SMITH, A., 1994 Somatic APC and K-ras codon 12 mutations in aberrant crypt foci from human colons. Cancer Research .

SPRADLING, A., D. DRUMMOND-BARBOSA, and T. KAI, 2001 Stem cells find their niche. Nature **414**: 98–104.

STONE, J. G., I. P. TOMLINSON, and R. S. HOULSTON, 2000 Optimising methods for determining RER status in colorectal cancers. Cancer Lett **149**: 15–20.

SU, L. K., K. W. KINZLER, B. VOGELSTEIN, A. C. PREISINGER, A. R. MOSER, C. LUONGO, K. A. GOULD, and W. F. DOVE, 1992 Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science **256**: 668–70.

TAKAYAMA, T., M. OHI, T. HAYASHI, and K. MIYANISHI, 2001 Analysis of K-ras, APC, and beta-Catenin in Aberrant Crypt Foci in Sporadic Adenoma, Cancer, and .... Gastroenterology .

TAYLOR, R. W., M. J. BARRON, G. M. BORTHWICK, A. GOSPEL, P. F. CHINNERY, D. C. SAMUELS, G. A. TAYLOR, S. M. PLUSA, S. J. NEEDHAM, L. C. GREAVES, T. B. L. KIRKWOOD, and D. M. TURNBULL, 2003 Mitochondrial DNA mutations in human colonic crypt stem cells. Journal of Clinical Investigation **112**: 1351–60.

THEISE, N. D., 2004 Perspective: stem cells react! Cell lineages as complex adaptive systems. Exp Hematol **32**: 25–7.

TOMLINSON, I., S. HALFORD, L. AALTONEN, N. HAWKINS, and R. WARD, 2002 Does MSI-low exist? J Pathol **197**: 6–13.

TOMLINSON, I. P., M. R. NOVELLI, and W. F. BODMER, 1996 The mutation rate and cancer. Proc Natl Acad Sci USA **93**: 14800–3.

TOTAFURNO, J., M. BJERKNES, and H. CHENG, 1987 The crypt cycle. Crypt and villus production in the adult intestinal epithelium. Biophys J **52**: 279–94.

TOTAFURNO, J., M. BJERKNES, and H. CHENG, 1988 Variation in crypt size and its influence on the analysis of epithelial cell proliferation in the intestinal crypt. Biophys J **54**: 845–58.

TSAO, J. L., Y. YATABE, R. SALOVAARA, H. J. JÄRVINEN, J. P. MECKLIN, L. A. AALTONEN, S. TAVARÉ, and D. SHIBATA, 2000 Genetic reconstruction of individual colorectal tumor histories. Proc Natl Acad Sci USA **97**: 1236–41.

TULINA, N. and E. MATUNIS, 2001 Control of stem cell self-renewal in Drosophila spermatogenesis by JAK-STAT signaling. Science **294**: 2546–9.

TUVESON, D. A., A. T. SHAW, N. A. WILLIS, D. P. SILVER, E. L. JACKSON, S. CHANG, K. L. MERCER, R. GROCHOW, H. HOCK, D. CROWLEY, S. R. HINGORANI, T. ZAKS, C. KING, M. A. JACOBETZ, L. WANG, R. T. BRONSON, S. H. ORKIN, R. A. DEPINHO, and T. JACKS, 2004 Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell **5**: 375–87.

UK OFFICE OF NATIONAL STATISTICS, Life expectancy. http://www.statistics.gov.uk/cci/nugget.asp?id=881.

VEIGL, M. L., L. KASTURI, J. OLECHNOWICZ, A. H. MA, J. D. LUTTERBAUGH, S. PERIYASAMY, G. M. LI, J. DRUMMOND, P. L. MODRICH, W. D. SEDWICK, and S. D. MARKOWITZ, 1998 Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci USA **95**: 8698–702.

VOGELSTEIN, B., E. R. FEARON, S. R. HAMILTON, S. E. KERN, A. C. PREISINGER, M. LEPPERT, Y. NAKAMURA, R. WHITE, A. M. SMITS, and J. L. BOS, 1988 Genetic alterations during colorectal-tumor development. N Engl J Med **319**: 525–32.

WALLENFANG, M. R., R. NAYAK, and S. DINARDO, 2006 Dynamics of the male germline stem cell population during aging of Drosophila melanogaster. Aging Cell **5**: 297–304.

WANG, Y., 2006 Epimorphin-/- mice have increased intestinal growth, decreased susceptibility to dextran sodium sulfate colitis, and impaired spermatogenesis. Journal of Clinical Investigation **116**: 1535–1546.

WASAN, H. S., H. S. PARK, K. C. LIU, N. K. MANDIR, A. WINNETT, P. SASIENI, W. F. BODMER, R. A. GOODLAD, and N. A. WRIGHT, 1998 APC in the regulation of intestinal crypt fission. J Pathol **185**: 246–55.

WATT, F. M. and B. L. HOGAN, 2000 Out of Eden: stem cells and their niches. Science **287**: 1427–30.

WEISSMAN, I. L., D. J. ANDERSON, and F. GAGE, 2001 Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. Annu Rev Cell Dev Biol **17**: 387–403.

WELLCOME TRUST SANGER INSTITUTE, Catologue of Somatic Mutations in Cancer. http://www.sanger.ac.uk/genetics/CGP/cosmic/.

WHITEHALL, V. L., M. D. WALSH, J. YOUNG, B. A. LEGGETT, and J. R. JASS, 2001 Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. Cancer Res **61**: 827–30.

WINTON, D. J., M. A. BLOUNT, and B. A. PONDER, 1988 A clonal marker induced by mutation in mouse intestinal epithelium. Nature **333**: 463–6.

WINTON, D. J. and B. A. PONDER, 1990 Stem-cell organization in mouse small intestine. Proc Biol Sci **241**: 13–8.

WOLFRAM RESEARCH, INC., 2005 *Mathematica Edition: Version 5.2*. Wolfram Research, Inc., Champaign, Illinois.

WONG, W.-M., N. MANDIR, R. A. GOODLAD, B. C. Y. WONG, S. B. GARCIA, S.-K. LAM, and N. A. WRIGHT, 2002 Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission. Gut **50**: 212–7.

WOOD, L. D., D. W. PARSONS, S. JONES, J. LIN, T. SJÖBLOM, R. J. LEARY, D. SHEN, S. M. BOCA, T. BARBER, J. PTAK, N. SILLIMAN, S. SZABO, Z. DEZSO, V. USTYANKSKY, T. NIKOLSKAYA, Y. NIKOLSKY, R. KARCHIN, P. A. WILSON, J. S. KAMINKER, Z. ZHANG, R. CROSHAW, J. WILLIS, D. DAWSON, M. SHIPITSIN, J. K. V. WILLSON, S. SUKUMAR, K. POLYAK, B. H. PARK, C. L. PETHIYAGODA, P. V. K. PANT, D. G. BALLINGER, A. B. SPARKS, J. HARTIGAN, D. R. SMITH, E. SUH, N. PAPADOPOULOS, P. BUCKHAULTS, S. D. MARKOWITZ, G. PARMIGIANI, K. W. KINZLER, V. E. VELCULESCU, and B. VOGELSTEIN, 2007 The genomic landscapes of human breast and colorectal cancers. Science **318**: 1108–13.

WRIGHT, C. M., O. F. DENT, R. C. NEWLAND, M. BARKER, P. H. CHAPUIS, E. L. BOKEY, J. P. YOUNG, B. A. LEGGETT, J. R. JASS, and G. A. MACDONALD, 2005 Low level microsatellite instability may be associated with reduced cancer specific survival in sporadic stage C colorectal carcinoma. Gut **54**: 103–8.

YAMADA, S., M. YASHIRO, K. MAEDA, Y. NISHIGUCHI, and K. HIRAKAWA, 2005 A novel high-specificity approach for colorectal neoplasia: Detection of K-ras2 oncogene mutation in normal mucosa. Int J Cancer **113**: 1015–21.

YAMASHITA, Y. M., D. L. JONES, and M. T. FULLER, 2003 Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. Science **301**: 1547–50.

YAMASHITA, Y. M., A. P. MAHOWALD, J. R. PERLIN, and M. T. FULLER, 2007 Asymmetric inheritance of mother versus daughter centrosome in stem cell division. Science **315**: 518–21.

YATABE, Y., S. TAVARÉ, and D. SHIBATA, 2001 Investigating stem cells in human colon by using methylation patterns. Proc Natl Acad Sci USA **98**: 10839–44.

YUAN, P., M. H. SUN, J. S. ZHANG, X. Z. ZHU, and D. R. SHI, 2001 APC and K-ras gene mutation in aberrant crypt foci of human colon. World J Gastroenterol **7**: 352–6.

ZHU, D., P. KEOHAVONG, S. D. FINKELSTEIN, P. SWALSKY, A. BAKKER, J. WEISSFELD, S. SRIVASTAVA, and T. L. WHITESIDE, 1997 K-ras gene mutations in normal colorectal tissues from K-ras mutation-positive colorectal cancer patients. Cancer Research **57**: 2485–92.

Chapter 8

Appendices

8.1 COX patch size data

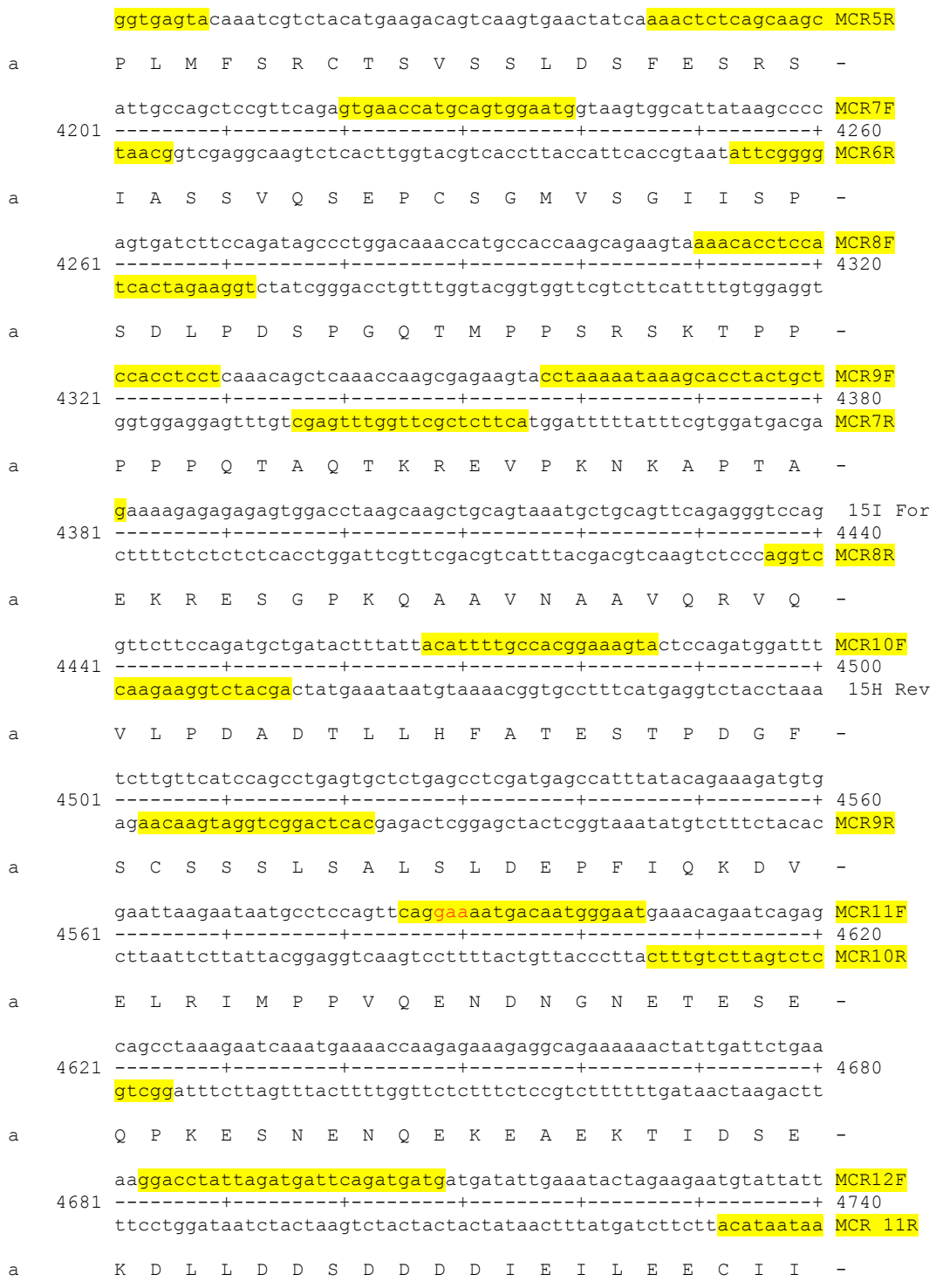
			Number of patches consisiting of X crypts																								
Sample ID	**Age**	**Total Crypts**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
1	25	163	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	25	321	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	32	192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	37	126	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	39	245	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	40	145	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	40	292	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	41	117	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	42	207	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	42	264	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	42	2702	20	6	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	43	197	10	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	43	198	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	43	212	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	45	131	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	46	272	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	47	174	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	47	174	13	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	47	245	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	48	250	11	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	49	197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	49	245	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	49	247	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	50	153	10	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	50	266	4	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	50	2878	20	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	51	197	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	51	242	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	52	167	11	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	52	250	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	55	113	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

			Number of patches consisiting of X crypts																								
Sample ID	**Age**	**Total Crypts**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
33	56	470	26	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	58	126	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	58	294	29	6	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	59	124	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	59	178	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	60	87	2	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	60	198	12	7	3	2	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	61	215	9	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	63	178	1	0	2	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
42	63	233	23	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	63	265	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	66	116	5	4	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	68	171	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	68	337	19	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	68	369	31	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
48	71	164	3	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	72	190	5	4	0	5	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	72	4101	52	11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	74	8792	164	65	20	9	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	76	194	7	6	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
53	80	6249	97	20	15	4	5	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54	82	10341	210	121	31	23	9	2	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
55	84	11821	89	51	40	16	8	1	3	2	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
56	75	1001	20	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	60	761	39	38	8	9	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	71	1917	35	17	22	7	5	1	3	0	1	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
59	67	1730	21	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	85	1293	24	12	2	0	2	2	1	0	0	0	2	2	0	1	0	0	0	0	0	0	0	0	0	0	0
61	84	1092	33	26	10	14	2	2	1	4	2	0	2	2	0	1	0	0	0	0	0	0	0	0	0	0	0
62	78	175	18	4	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

8.2 APC sequence

The wild-type sequence of the mRNA of the APC mutator cluster region is displayed below. Second round APC primers (identical to the APC SSCP primers) are highlighted in yellow.

```
      tcatctggacaaagcagtaaaaccgaacatatgtcttcaagcagtgagaatacgtccaca  MCR1F
 3601 ---------+---------+---------+---------+---------+---------+ 3660
      agtagacctgtttcgtcattttggcttgtatacagaagttcgtcactcttatgcaggtgt

a        S  S  G  Q  S  S  K  T  E  H  M  S  S  S  S  E  N  T  S  T   -

      ccttcatctaatgccaagaggcagaatcagctccatccaagttctgcacagagtagaagt
 3661 ---------+---------+---------+---------+---------+---------+ 3720
      ggaagtagattacggttctccgtcttagtcgaggtaggttcaagacgtgtctcatcttca

a        P  S  S  N  A  K  R  Q  N  Q  L  H  P  S  S  A  Q  S  R  S   -

      ggtcagcctcaaaaggctgccacttgcaaagtttcttctattaaccaagaaacaatacag  15G MCR2F
 3721 ---------+---------+---------+---------+---------+---------+ 3780
      ccagtcggagttttccgacggtgaacgtttcaaagaagataattggttctttgttatgtc MCR1R

a        G  Q  P  Q  K  A  A  T  C  K  V  S  S  I  N  Q  E  T  I  Q   -

      acttattgtgtagaagatactccaatatgtttttcaagatgtagttcattatcatctttg MCR3F
 3781 ---------+---------+---------+---------+---------+---------+ 3840
      tgaataacacatcttctatgaggttatacaaaaagttctacatcaagtaatagtagaaac  15F Rev

a        T  Y  C  V  E  D  T  P  I  C  F  S  R  C  S  S  L  S  S  L   -

      tcatcagctgaagatgaaataggatgtaatcagacgacacaggaagcagattctgctaat
 3841 ---------+---------+---------+---------+---------+---------+ 3900
      agtagtcgacttctactttatcctacattagtctgctgtgtccttcgtctaagacgatta MCR2R

a        S  S  A  E  D  E  I  G  C  N  Q  T  T  Q  E  A  D  S  A  N   -

      accctgcaaatagcagaaataaaagaaaagattggaactaggtcagctgaagatcctgtg MCR4F
 3901 ---------+---------+---------+---------+---------+---------+ 3960
      tgggacgtttatcgtctttatttttcttttctaaccttgatccagtcgacttctaggacac

a        T  L  Q  I  A  E  I  K  E  K  I  G  T  R  S  A  E  D  P  V   -

      agcgaagttccagcagtgtcacagcaccctagaaccaaatccagcagactgcagggttct
 3961 ---------+---------+---------+---------+---------+---------+ 4020
      tcgcttcaaggtcgtcacagtgtcgtgggatcttggtttaggtcgtctgacgtcccaaga MCR3R

a        S  E  V  P  A  V  S  Q  H  P  R  T  K  S  S  R  L  Q  G  S   -

      agtttatcttcagaatcagccaggcacaaagctgttgaattttcttcaggagcgaaatct  15H For MCR5F
 4021 ---------+---------+---------+---------+---------+---------+ 4080
      tcaaatagaagtcttagtcggtccgtgtttcgacaacttaaaagaagtcctcgctttaga

a        S  L  S  S  E  S  A  R  H  K  A  V  E  F  S  S  G  A  K  S   -

      ccctccaaaagtggtgctcagacacccaaaagtccacctgaacactatgttcaggagacc MCR6F
 4081 ---------+---------+---------+---------+---------+---------+ 4140
      gggaggttttcaccacgagtctgtgggttttcaggtggacttgtgatacaagtcctctgg  15G Rev MCR4R

a        P  S  K  S  G  A  Q  T  P  K  S  P  P  E  H  Y  V  Q  E  T   -

      ccactcatgtttagcagatgtacttctgtcagttcacttgatagttttgagagtcgttcg
 4141 ---------+---------+---------+---------+---------+---------+ 4200
```


ggtgagtacaaatcgtctacatgaagacagtcaagtgaactatcaaaactctcagcaagc MCR5R
a P L M F S R C T S V S S L D S F E S R S -
attgccagctccgttcagagtgaaccatgcagtggaatggtaagtggcattataagcccc MCR7F
4201 ---------+---------+---------+---------+---------+---------+ 4260
taacggtcgaggcaagtctcacttggtacgtcaccttaccattcaccgtaatattcgggg MCR6R
a I A S S V Q S E P C S G M V S G I I S P -
agtgatcttccagatagccctggacaaaccatgccaccaagcagaagtaaaacacctcca MCR8F
4261 ---------+---------+---------+---------+---------+---------+ 4320
tcactagaaggtctatcgggacctgtttggtacggtggttcgtcttcattttgtggaggt
a S D L P D S P G Q T M P P S R S K T P P -
ccacctcctcaaacagctcaaaccaagcgagaagtacctaaaaataaagcacctactgct MCR9F
4321 ---------+---------+---------+---------+---------+---------+ 4380
ggtggaggagtttgtcgagtttggttcgctcttcatggatttttatttcgtggatgacga MCR7R
a P P P Q T A Q T K R E V P K N K A P T A -
gaaaagagagagagtggacctaagcaagctgcagtaaatgctgcagttcagagggtccag 15I For
4381 ---------+---------+---------+---------+---------+---------+ 4440
cttttctctctctcacctggattcgttcgacgtcatttacgacgtcaagtctcccaggtc MCR8R
a E K R E S G P K Q A A V N A A V Q R V Q -
gttcttccagatgctgatactttattacattttgccacggaaagtactccagatggattt MCR10F
4441 ---------+---------+---------+---------+---------+---------+ 4500
caagaaggtctacgactatgaaataatgtaaaacggtgcctttcatgaggtctacctaaa 15H Rev
a V L P D A D T L L H F A T E S T P D G F -
tcttgttcatccagcctgagtgctctgagcctcgatgagccatttatacagaaagatgtg
4501 ---------+---------+---------+---------+---------+---------+ 4560
agaacaagtaggtcggactcacgagactcggagctactcggtaaatatgtctttctacac MCR9R
a S C S S S L S A L S L D E P F I Q K D V -
gaattaagaataatgcctccagttcaggaaaatgacaatgggaatgaaacagaatcagag MCR11F
4561 ---------+---------+---------+---------+---------+---------+ 4620
cttaattcttattacggaggtcaagtccttttactgttacccttactttgtcttagtctc MCR10R
a E L R I M P P V Q E N D N G N E T E S E -
cagcctaaagaatcaaatgaaaaccaagagaaagaggcagaaaaaactattgattctgaa
4621 ---------+---------+---------+---------+---------+---------+ 4680
gtcggatttcttagtttacttttggttctctttctccgtctttttgataactaagactt
a Q P K E S N E N Q E K E A E K T I D S E -
aaggacctattagatgattcagatgatgatgatattgaaatactagaagaatgtattatt MCR12F
4681 ---------+---------+---------+---------+---------+---------+ 4740
ttcctggataatctactaagtctactactactataactttatgatcttcttacataataa MCR 11R
a K D L L D D S D D D D I E I L E E C I I -

```
      tctgccatgccaacaaagtcatcacgtaaagcaaaaaagccagcccagactgcttcaaaa  15J For
 4741 ---------+---------+---------+---------+---------+---------+ 4800
      agacggtacggtgtttcagtagtgcatttcgtttttcggtcgggtctgacgaagtttt

a     S  A  M  P  T  K  S  S  R  K  A  K  K  P  A  Q  T  A  S  K   -

      ttacctccacctgtggcaaggaaaccaagtcagctgcctgtgtacaaacttctaccatca
 4801 ---------+---------+---------+---------+---------+---------+ 4860
      aatggaggtggacaccgttcctttggttcagtcgacggacacatgtttgaagatggtagt MCR12R

a     L  P  P  P  V  A  R  K  P  S  Q  L  P  V  Y  K  L  L  P  S   -
```

8.3 K-ras sequence

The wild-type sequence of the mRNA of the K-ras gene is displayed below. Codons 12 and 13, the two most frequently mutated codons in colorectal cancer are highlighted in yellow.

```
atg act gaa tat aaa ctt gtg gta gtt gga gct ggt ggc gta ggc aag agt gcc ttg acg   60
  M   T   E   Y   K   L   V   V   V   G   A   G   G   V   G   K   S   A   L   T   20

ata cag cta att cag aat cat ttt gtg gac gaa tat gat cca aca ata gag gat tcc tac   120
  I   Q   L   I   Q   N   H   F   V   D   E   Y   D   P   T   I   E   D   S   Y   40

agg aag caa gta gta att gat gga gaa acc tgt ctc ttg gat att ctc gac aca gca ggt   180
  R   K   Q   V   V   I   D   G   E   T   C   L   L   D   I   L   D   T   A   G   60

caa gag gag tac agt gca atg agg gac cag tac atg agg act ggg gag ggc ttt ctt tgt   240
  Q   E   E   Y   S   A   M   R   D   Q   Y   M   R   T   G   E   G   F   L   C   80

gta ttt gcc ata aat aat act aaa tca ttt gaa gat att cac cat tat aga gaa caa att   300
  V   F   A   I   N   N   T   K   S   F   E   D   I   H   H   Y   R   E   Q   I   100

aaa aga gtt aag gac tct gaa gat gta cct atg gtc cta gta gga aat aaa tgt gat ttg   360
  K   R   V   K   D   S   E   D   V   P   M   V   L   V   G   N   K   C   D   L   120

cct tct aga aca gta gac aca aaa cag gct cag gac tta gca aga agt tat gga att cct   420
  P   S   R   T   V   D   T   K   Q   A   Q   D   L   A   R   S   Y   G   I   P   140

ttt att gaa aca tca gca aag aca aga cag ggt gtt gat gat gcc ttc tat aca tta gtt   480
  F   I   E   T   S   A   K   T   R   Q   G   V   D   D   A   F   Y   T   L   V   160

cga gaa att cga aaa cat aaa gaa aag atg agc aaa gat ggt aaa aag aag aaa aag aag   540
  R   E   I   R   K   H   K   E   K   M   S   K   D   G   K   K   K   K   K   K   180

tca aag aca aag tgt gta att atg taa                                             567
  S   K   T   K   C   V   I   M   *                                             189
```

8.4 Microsatellite genotyping results

			D18S58					D14S144				
P,tientID	,ge	MSIH	g1	g2	s1	s2	s3	s4	g1	g2	s1	s2	s3	s4
1	23	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
2	50	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
3	53	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
4	54	0	0	0	0	ns	ns	ns	0	1	-6,-5,0,1	-3,-2,0,1	-2,0,1	ns
5	55	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
6	56	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
7	57	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
8	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
9	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
10	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
11	59	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
12	60	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H1	63	1	0	4	-4,-3,-2,-1,0	-4,-3,-2,-1,0	-4,-3,-2,-1,0	-4,-3,-2,-1,0	0	1	-2,-1,0	-2,-1,0	-2,-1,0	-2,-1,0
13	63	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
14	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
15	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
16	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
17	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
18	64	0	0	1	0,1	0,1	0,1	-1,0,1,2,3	0	0	0	ns	ns	ns
19	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
20	65	0	0	1	0,1	-2,-1,0,1,2	0,1	-1,0,1,2	0	1	-3,-2,0,1	0,1	0,1	0,1
21	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
22	65	0	0	0	0	ns	ns	ns	0	1	0,1,2	0,1	0,1,2	0,1
23	66	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H2	67	1	0	3	-2,-1,0,1,2,3	-2,-1,0,1,2,3	-2,-1,0,1,2,3	-2,-1,0,1,2,3	0	2	-3,0,2	-3,0,2	-3,0,2	-3,-2,-1,0,2
24	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
25	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns

					D10S197						D15S659			
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
1	23	0	0	0	0	ns	ns	ns	0	0	failed	failed	failed	failed
2	50	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
3	53	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
4	54	0	0	0	0	ns	ns	ns	0	4	0,2,4,5	0,2,4,5	-1,0,1,2,3,4	0,1,2,3,4,6
5	55	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
6	56	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
7	57	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
8	58	0	0	1	0,1	0,1	-3,-2,0,1	0,1	0	0	0	ns	ns	ns
9	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
10	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
11	59	0	0	0	failed	failed	failed	failed	0	0	0	ns	ns	ns
12	60	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H1	63	1	0	1	-4,-3,-2,0,1	-4,-3,-2,0,1	-4,-3,-1,0,1	-4,-3,-1,0,1	0	0	failed	failed	failed	failed
13	63	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
14	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
15	64	0	0	0	0	ns	ns	ns	0	1	0,1,2	0,1,2	0,1,2	0,1
16	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
17	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
18	64	0	0	1	0,1,2	0,1	0,1,2	0,1	0	0	failed	failed	failed	failed
19	65	0	0	1	-2,0,1	-3,0,1	0,1	0,1	0	0	0	ns	ns	ns
20	65	0	0	0	0	ns	ns	ns	0	1	0,1	0,1,3	0,1,2,3	0,1,3
21	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
22	65	0	0	0	0	ns	ns	ns	0	2	-1,0,1,2,3	-1,0,1,2,3	-1,0,1,2,3	0,1,2,3,4
23	66	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H2	67	1	0	1	-3,-2,0,1,2,3	-3,-2,0,1,2,3	-3,-2,0,1,2,3	0,1,2,3,4,5	0	0	0	ns	ns	ns
24	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
25	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns

			D13S175						D17S250					
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
1	23	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
2	50	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
3	53	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
4	54	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
5	55	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
6	56	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
7	57	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
8	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
9	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
10	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
11	59	0	0	0	0	ns	ns	ns	0	2	0,2	-6,-4,0,2	0,2	0,2
12	60	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H1	63	1	0	1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1	0	0	failed	failed	failed	failed
13	63	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
14	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
15	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
16	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
17	64	0	0	0	0	ns	ns	ns	0	0	failed	failed	failed	failed
18	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
19	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
20	65	0	0	0	0	ns	ns	ns	0	0	failed	failed	failed	failed
21	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
22	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
23	66	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H2	67	1	0	1	-5,-4,-3,-2,-1,0,1	-5,-4,-3,-2,-1,0,1	-5,-4,-3,-2,-1,0,1	-5,-4,-3,-2,-1,0,1	0	0	failed	failed	failed	failed
24	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
25	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns

			D2S123						D5S346					
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
1	23	0	0	1	0,1,2	0,1,2	0,1,2	0,1,2	0	0	0	ns	ns	ns
2	50	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
3	53	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
4	54	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
5	55	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
6	56	0	0	0	0	ns	ns	ns	0	3	0,3	0,3	-3,0,3	0,3
7	57	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
8	58	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
9	58	0	0	1	0,1	0,1	0,1	0,1,5,6,7	0	0	0	ns	ns	ns
10	58	0	0	2	0,2,6,7	0,2	-2,0,2	0,2	0	0	0	ns	ns	ns
11	59	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
12	60	0	0	6	0,1,6	0,6	0,6	0,6	0	0	0	ns	ns	ns
H1	63	1	0	4	-5,-4,-3,0,3,4	-5,-4,0,3,4	-5,0,2,4	-3,0,4	0	1	0,1,2	0,1,2	0,1,2	0,1,2
13	63	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
14	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
15	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
16	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
17	64	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
18	64	0	0	1	-1,0,1,4,5	0,1	0,1	-1,0,1,4,5	0	0	0	ns	ns	ns
19	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
20	65	0	0	0	failed	failed	failed	failed	0	0	0	ns	ns	ns
21	65	0	0	6	0,6	0,1,2,6	0,1,2,6	0,6	0	0	0	ns	ns	ns
22	65	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
23	66	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H2	67	1	0	0	failed	failed	failed	failed	0	2	-4,-3,0,2	-4,-3,0,2	-4,-3,0,2	-4,-3,0,2
24	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
25	67	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns

P,tientID	,ge	MSIH	g1	g2	D8S87 s1	s2	s3	s4
1	23	0	0	0	0	ns	ns	ns
2	50	0	0	0	0	ns	ns	ns
3	53	0	0	0	0	ns	ns	ns
4	54	0	0	0	0	ns	ns	ns
5	55	0	0	0	0	ns	ns	ns
6	56	0	0	0	0	ns	ns	ns
7	57	0	0	0	0	ns	ns	ns
8	58	0	0	0	0	ns	ns	ns
9	58	0	0	0	0	ns	ns	ns
10	58	0	0	0	0	ns	ns	ns
11	59	0	0	0	0	ns	ns	ns
12	60	0	0	0	0	ns	ns	ns
H1	63	1	0	1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1	-4,-3,-2,-1,0,1
13	63	0	0	0	0	ns	ns	ns
14	64	0	0	0	0	ns	ns	ns
15	64	0	0	0	0	ns	ns	ns
16	64	0	0	0	0	ns	ns	ns
17	64	0	0	0	0	ns	ns	ns
18	64	0	0	0	0	ns	ns	ns
19	65	0	0	1	0,1	0,1	0,1	0,1,2
20	65	0	0	0	failed	failed	failed	failed
21	65	0	0	0	0	ns	ns	ns
22	65	0	0	0	0	ns	ns	ns
23	66	0	0	0	0	ns	ns	ns
H2	67	1	0	2	0,2,3	0,2,3	0,2,3	0,2,3
24	67	0	0	0	0	ns	ns	ns
25	67	0	0	0	0	ns	ns	ns
H3	69	1	0	1	0,1,2,3,4,5	0,1,2,3,4,5	0,1,2,3,4,5	0,1,2,3,4,5

					D18S58						D14S144			
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
H3	69	1	0	0	0	ns	ns	ns	0	1	0,1,2,3,4,5,6,7	0,1,2,3,4,5,6,7	0,1,2,3,4,5,6,7	0,1,2,3,4,5,6,7
26	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
27	70	0	0	0	0	ns	ns	ns	0	1	0,1,2	0,1,2	0,1,2	0,1,2
28	71	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
29	71	0	0	2	-3,0,2,3	-6,-1,0,2	-2,0,2	-1,0,1,2	0	0	0	ns	ns	ns
30	72	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
31	73	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
32	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
33	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H4	75	1	0	1	-2,-1,0,1,2,3	-2,-1,0,1,2,3	-2,1,0	-2,1,0,1	0	1	-3,-2,0,1	-3,-2,0,1	-3,-2,0,1	-3,-2,0,1
34	78	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
35	78	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
36	79	0	0	3	-3,-2,-1,0,2	-2,-1,0,3	-2,-1,0,3	-1,1,4	0	0	0	ns	ns	ns
37	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
38	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
39	81	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
40	83	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
41	84	0	0	1	0,1	0,1	0,1	0,1	0	0	0	ns	ns	ns
42	86	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H5	103	1	0	3	-5,0,3,6	-5,0,3,6	-5,0,3,6	-5,0,3,6	0	0	0	ns	ns	ns

			D10S197					D15S659				
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
H3	69	1	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
26	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
27	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
28	71	0	0	0	failed	failed	failed	failed	0	1	0,1,2,3	0,1,2	0,1,3	0,1
29	71	0	0	0	failed	failed	failed	failed	0	1	0,1,2,3,4	0,1,2,4	0,1,3,4	0,1,3,4,6
30	72	0	0	0	0	ns	ns	ns	0	1	0,1,2,3	0,1,2,3	0,1,2,3	0,1,2,3
31	73	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
32	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
33	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H4	75	1	0	1	-3,-2,0,1	-3,-2,0,1	-3,-2,0,1	-3,-2,0,1	0	0	0	ns	ns	ns
34	78	0	0	0	0	ns	ns	ns	0	1	0,1	0,1	-2,0,1	0,1
35	78	0	0	1	0,1	0,1	-2,0,1	0,1	0	0	0	ns	ns	ns
36	79	0	0	0	failed	failed	failed	failed	0	0	failed	failed	failed	failed
37	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
38	80	0	0	0	0	ns	ns	ns	0	1	0,1	-5,-4,1	-4,1	0,1
39	81	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
40	83	0	0	0	0	ns	ns	ns	0	5	0,2,4,5	1,2,5	0,2,4,5	-4,0,1,2,4,6
41	84	0	0	0	0	ns	ns	ns	0	1	0,1,2,5	0,1	0,1,2,5	0,1
42	86	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H5	103	1	0	1	-6,-5,-4,-3,-2,0,1	-6,-5,-4,-3,-2,0,1	-6,-5,-4,-3,-2,0,1	-6,-5,-4,-3,-2,0,1	0	1	-1,0,3,5,6	-1,0,2,4	-1,0,1	-1,0,1,3,5,6

			D13S175						D17S250					
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
H3	69	1	0	0	0	ns	ns	ns	0	1	-1,0,1	-1,0,1	-1,0,1	-1,0
26	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
27	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
28	71	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
29	71	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
30	72	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
31	73	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
32	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
33	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H4	75	1	0	2	0,2	0,2	0,2	0,2	0	0	failed	failed	failed	failed
34	78	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
35	78	0	0	0	0	ns	ns	ns	0	2	0,2	-1,0,2	-2,-1,0,2	0,2
36	79	0	0	1	-1,0,1,2	-1,0,1,2	-1,0,1,2	-1,0,1,2	0	0	failed	failed	failed	failed
37	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
38	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
39	81	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
40	83	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
41	84	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
42	86	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H5	103	1	0	3	0,2,3,5	0,1,2,3,5	0,1,2,3	0,1,2,3	0	2	-2,-1,0,2,4	-2,-1,0,2,4	-2,-1,0,2,4	-2,-1,0,2,4

			D2S123						D5S346					
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
H3	69	1	0	1	0,1,2	0,1,2	0,1,2	0,1,6,7	0	4	0,4,5	0,4,5	0,4,5	0,4,5
26	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
27	70	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
28	71	0	0	0	failed	failed	failed	failed	0	0	failed	failed	failed	failed
29	71	0	0	0	failed	failed	failed	failed	0	0	0	ns	ns	ns
30	72	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
31	73	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
32	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
33	75	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H4	75	1	0	1	-2,-1,0,1,2	-2,-1,0,1	-2,-1,0,1	-2,-1,0,1	0	0	0	ns	ns	ns
34	78	0	0	0	failed	failed	failed	failed	0	0	0	ns	ns	ns
35	78	0	0	0	0	ns	ns	ns	0	1	0,1	0,1	0,1,2	0,1
36	79	0	0	0	failed	failed	failed	failed	0	0	0	ns	ns	ns
37	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
38	80	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
39	81	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
40	83	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
41	84	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
42	86	0	0	0	0	ns	ns	ns	0	0	0	ns	ns	ns
H5	103	1	0	1	-5,-2,-1,0,7,8	-6,-5,0,1	-5,-4,-1,0	-5,-4,-3,-2,0,1	0	4	01,2,3,4	01,2,3,4	01,2,3,4	01,2,3,4

					D8S87			
P,tientID	**,ge**	**MSIH**	**g1**	**g2**	**s1**	**s2**	**s3**	**s4**
26	70	0	0	0	0	ns	ns	ns
27	70	0	0	0	0	ns	ns	ns
28	71	0	0	0	failed	failed	failed	failed
29	71	0	0	0	failed	failed	failed	failed
30	72	0	0	0	0	ns	ns	ns
31	73	0	0	0	0	ns	ns	ns
32	75	0	0	0	0	ns	ns	ns
33	75	0	0	0	0	ns	ns	ns
H4	75	1	0	1	-2,-1,1	-2,-1,1	-1,0,1	-1,0,1
34	78	0	0	0	0	ns	ns	ns
35	78	0	0	0	0	ns	ns	ns
36	79	0	0	0	0	ns	ns	ns
37	80	0	0	0	0	ns	ns	ns
38	80	0	0	0	0	ns	ns	ns
39	81	0	0	0	0	ns	ns	ns
40	83	0	0	0	0	ns	ns	ns
41	84	0	0	0	0	ns	ns	ns
42	86	0	0	0	0	ns	ns	ns
H5	103	1	0	1	-3,-2,-1,0,1	-3,-2,-1,0,1	-3,-2,-1,0,1	-3,-2,-1,0,1